

Realizzazione della “Camera dei Diamanti” e verifica delle prestazioni di spazializzazione del suono

Jacopo Grassi^{a*} | Nicola Prodi^a

^a Dipartimento di Ingegneria,
Università degli Studi di Ferrara,
Via Giuseppe Saragat, 1, 44122 Ferrara

* Autore di riferimento:
jacopo.grassi@unife.it

Ricevuto: 9/1/2025

Accettato: 3/3/2025

DOI: 10.3280/ria1-2025oa19121

ISSN: 2385-2615

La “Camera dei Diamanti” è una piattaforma per realtà virtuale sviluppata presso l’Università di Ferrara. Le principali applicazioni spaziano dalla medicina alla psicologia e all’industria. Ad esempio: ricerca su protesi acustiche, intelligibilità della parola, sound quality dei prodotti, formazione e sicurezza. È illustrato il processo di realizzazione del sistema formato da una cabina audiometrica con al suo interno 41 altoparlanti (di cui un subwoofer) e un visore per realtà virtuale. Questa strumentazione provvede alla riproduzione di ambienti virtuali realistici; la parte audio è ottenuta grazie alla tecnologia Ambisonics, ormai diventata uno standard del settore; quella video è stata sviluppata con Unity.

È stata ideata una metodologia oggettiva che sfrutta array microfonic per stimare la bontà della spazializzazione del suono, utile a verificare il funzionamento del sistema e conoscerne i limiti. I risultati ottenuti per la “Camera dei Diamanti” sono particolarmente incoraggianti poiché l’errore di riproduzione di una sorgente sonora singola virtuale o reale si aggira attorno a 1°-2°, valori simili alla capacità umana di distinguere due suoni come provenienti da direzioni diverse in condizioni ottimali.

Parole chiave: Ambisonics, spazializzazione del suono, localizzazione, realtà virtuale, array microfonic, minimum audible angle

Realization of the “Camera dei Diamanti” and performance evaluation of the sound spatialization

“Camera dei Diamanti” is the virtual reality lab of the University of Ferrara. The main applications range from medicine and psychology to industry. Practical examples are acoustical prosthesis research, speech intelligibility, sound quality of industrial products, education and safety training. This work shows the physical realization of the hardware, so how 41 loudspeakers (including a subwoofer), an acoustic treatment and virtual reality glasses were added to an audiometric chamber. These make it possible to reproduce realistic virtual reality scenarios; the audio playback is achieved using Ambisonics technology, that is a standard nowadays for those labs; the video reproduction is managed by Unity.

An objective methodology, based on microphone arrays, has been developed to evaluate the performance of spatialization of the sound in order to be aware of the hardware’s limits.

The results obtained for the “Camera dei Diamanti” are encouragingly positive since the reproduction error evaluated for single source, both real and virtual ones, is around 1°-2°. These values are comparable to the human capability to distinguish two sounds coming from different directions under optimal conditions.

Keywords: Ambisonics, sound spatialization, localization, virtual reality, microphone arrays, minimum audible angle

1 | Introduzione

1.1 | Laboratorio di realtà virtuale

I laboratori per la realtà virtuale sono una delle risorse ad oggi più avanzate per la riproduzione di scenari virtuali audio-video. Costano di un array di altoparlanti e di un sistema di riproduzione video come visori per la realtà virtuale o proiettori. Strutture del genere sono spesso collocate in camere anecoiche per avere una restituzione audio molto fedele e ben controllabile.

Tali laboratori hanno suscitato particolare interesse nell’ambiente internazionale della ricerca, soprattutto in psi-

coacustica e nelle scienze per l’udito. Ad esempio in ambito audiologico è possibile ricreare ambienti acustici complessi [1] che non sono facilmente riproducibili con le tecniche di laboratorio più classiche. Presentando il Virtual Reality Lab dell’Università di Oldenburg, Hohmann et al. [2] mostrano come sia importante l’utilizzo di tali strutture per la valutazione e la progettazione di dispositivi innovativi per i deficit uditivi perché è rilevante ridurre al minimo la discrepanza tra sperimentazioni in laboratorio e l’esperienza nella vita reale [3-5]. Con i laboratori per la realtà virtuale si possono proporre efficacemente l’insegnamento e la riabilitazione alla distinzione di diversi stimoli sonori provenienti da direzioni differenti, in

condizioni controllate, ripetibili e molto simili alla realtà quotidiana [2].

I laboratori in questione sono dotati di strumentazione avanzata, ad esempio quello dell'Università di Oldenburg gestisce la parte video con uno schermo cilindrico (oppure con un visore per realtà virtuale HTC Vive [6]) e la parte audio con 28 altoparlanti nascosti dietro allo schermo cilindrico stesso. Viene anche utilizzato il sistema TASCAR [7] che permette di scegliere il tipo di rendering audio. Alla Università Tecnica della Danimarca (DTU) è presente AVIL [8] che consta di un sistema formato da 72 altoparlanti e un visore HTC VIVE PRO. In Italia sono presenti altre due realtà simili: "Stanza di Matilde" al IRCCS materno infantile Burlo Garofolo e "Audio Space Lab" al Politecnico di Torino. In particolare quest'ultima [9] è una stanza equipaggiata con 18 altoparlanti e un visore Meta Quest 2. Altri esempi sono SCaLAr dell'Università RWTH di Aachen [10] e ALF [11] della Base Air Force Wright-Patterson che sono formate rispettivamente da array di 64 e 277 altoparlanti.

La "Camera dei Diamanti" è un laboratorio per la realtà virtuale dell'Università di Ferrara. Il nome richiama uno dei simboli della città, il Palazzo dei Diamanti, le cui facciate sono completamente ricoperte da cunei, vedi Fig. 1. All'interno della camera è presente un pattern simile ottenuto grazie ad un trattamento acustico formato da coni di melammina, vedi Fig. 5.



Fig. 1 – Palazzo dei Diamanti
Palazzo dei Diamanti

Tieghi Maurizio, CC BY-SA 4.0, https://commons.wikimedia.org/wiki/File:Ferrara_palazzo_diamanti.jpg

È uno spazio adibito alla riproduzione di scenari virtuali visivo-uditivi tramite un array di 41 altoparlanti e un visore per la realtà virtuale Meta Quest 3. La tecnologia sfruttata per la riproduzione audio è Ambisonics (vedi paragrafo 1.2): l'utilizzo di semplici cuffie avrebbe limitato la percezione, ad esempio l'esternalizzazione del suono sarebbe più difficoltosa, ovvero il soggetto che indossa cuffie è più portato a "sentire il suono dentro la testa"; ma avrebbe limitato anche applicazioni di carattere medico come la valutazione delle performance di localizzazione del suono o di intelligibilità della parola di soggetti con protesi acustiche.

Il punto di forza, che distingue le indagini svolte all'interno di un laboratorio di realtà virtuale rispetto a quelle eseguite in ambienti più "tradizionali", è l'alto livello di sensazione di immersione all'interno della scena in cui ci si trova. La percezione della qualità di un prodotto industriale è fortemente

dipendente dalla presentazione di un'informazione visivo-uditiva, dalla sua qualità e dalla capacità di questa di essere trasmessa in modo fedele e immersivo al soggetto interessato. Lo stesso ragionamento è valido per le applicazioni inerenti agli addestramenti di sicurezza in diversi ambienti.

L'obiettivo dello studio è presentare come è stata realizzata la Camera dei Diamanti e formulare una metodologia di stima delle prestazioni di spazializzazione del suono che sia semplice ed efficace. Viene proposta, inoltre, l'applicazione della metodologia stessa alla "Camera dei Diamanti".

1.2 | Ambisonics

Ambisonics è una tecnologia sviluppata per la registrazione e la riproduzione spazializzata del suono, che può operare in maniera indipendente dalla configurazione degli altoparlanti che eseguono il playback, a patto che questi siano disposti su una superficie quasi sferica [12]. Invece, l'acquisizione del segnale avviene attraverso array microfonici aventi molte capsule (tipicamente 19, 32 o 64) disposte su di una superficie sferica.

Zotter e Frank [12] mostrano come la tecnologia sia basata sulla descrizione del campo sonoro attraverso armoniche sferiche, vedi Fig. 2. Più elevato è il numero di armoniche sferiche disponibili, più cresce la precisione con cui avvengono la registrazione e la riproduzione in termini di spazialità. L'ordine Ambisonics, graficamente, è il "numero di riga del triangolo" in Fig. 2 partendo dallo 0 della punta in alto (ordine 0 significa avere un'armonica sola).

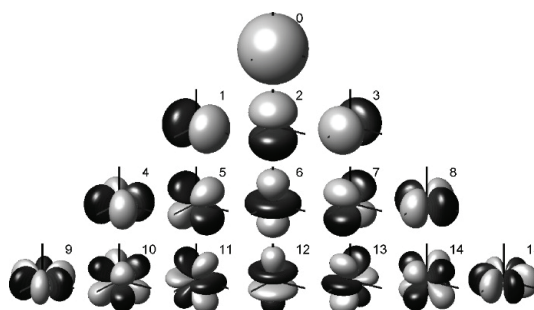


Fig. 2 – Armoniche sferiche fino al terzo ordine indicizzate con il numero di canale Ambisonics ACN (da Zotter e Frank [12])
Spherical harmonics up to the third order indexed with the Ambi-sonics channel number ACN (from Zotter and Frank [12])

Tecnicamente servirebbe un canale per ogni armonica sferica (si definisce Ambisonic Channel Number come $ACN = (O + 1)^2$, con O numero dell'ordine), dunque guardando la Fig. 2 ne servirebbero 16. Ciò significa che registrare un segnale al terzo ordine Ambisonics significa avere un array microfonico da almeno 16 capsule microfoniche orientate in modo diverso. Parlando invece di riproduzione, si può avere un playback del terzo ordine anche con un numero inferiore di altoparlanti, questo grazie al decoder ALLRAD, sviluppato da Zotter e Frank [13], che include nell'elaborazione del segnale il metodo di panning VBAP (vector-base amplitude panning),

ideato da Pulkki [14], ovvero un sistema di restituzione che permette il posizionamento delle sorgenti virtuali nello spazio.

Nel caso della “Camera dei Diamanti”, con 41 altoparlanti, sono state riprodotte tracce anche del settimo ordine. Per avere, invece, un settimo ordine “puro” sarebbero necessari 64 altoparlanti.

1.3 | Capacità di localizzazione di sorgenti singole

Per valutare le performances di spazializzazione e comprendere se l'errore di riproduzione, in termini di posizionamento dei suoni, è accettabile, bisogna aver ben presente i limiti dell'uomo nel percepire la direzione di un suono e prendere questi valori come riferimento.

Viene introdotto un sistema di riferimento sferico centrato nella testa dell'ascoltatore; la posizione, ad esempio di una sorgente, viene definita con gli angoli di azimut ed elevazione, vedi Fig. 3. Il primo descrive l'angolo che spazza il piano “orizzontale” che passa per gli occhi dell'ascoltatore con riferimento di 0° nella direzione frontale; alla destra dell'ascoltatore assume valori negativi e alla sinistra positivi, i valori sono compresi tra -180° e 180°. Il secondo angolo spazza sul piano, detto “mediano”, che taglia verticalmente la testa dell'ascoltatore; con riferimento alla direzione frontale si hanno 0°, con valori compresi tra 90° e -90°, sopra l'orizzontale sono positivi e sotto negativi. Il raggio della sfera è considerato costante e unitario.

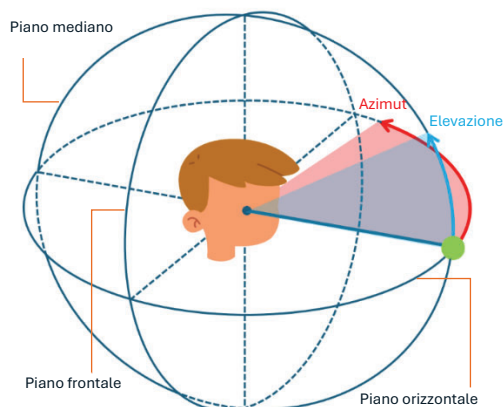


Fig. 3 – Sistema di riferimento (azimut in rosso ed elevazione in blu)
Reference system (azimuth in red and elevation in blue)

Blauert [15] definisce la localizzazione come “la legge o la regola in base alla quale la posizione di un evento uditivo è correlata a uno o più attributi specifici di un evento sonoro, o di un altro evento che è in qualche modo correlato all'evento uditivo.” Dove l'evento sonoro identifica il fenomeno fisico di propagazione dell'onda, mentre l'evento uditivo rappresenta la percezione di esso. Poi si definisce [15] anche la sfocatura di localizzazione come “la più piccola variazione di uno o più attributi specifici di un evento sonoro o di un altro evento correlato a un evento uditivo, sufficiente a produrre una va-

riazione nella posizione dell'evento uditivo. La sfocatura di localizzazione è una proprietà della localizzazione”. Blauert [15] continua paragonando la localizzazione a quella funzione che lega i punti del campo spaziale reale a quello della percezione, allora la sfocatura sarà la caratteristica che quantifica la “risoluzione” della spazialità percepita, che è minore rispetto a quella reale. La percezione della posizione di una sorgente singola dipende da molteplici fattori come la frequenza, la direzione di provenienza e la tipologia di segnale. In base alla combinazione dei fattori, possono insorgere anche illusioni uditive che posizionano la sorgente percepita in un punto completamente diverso da quello reale.

Carlini et al. [16] rimarcano come sia necessario aver presente che la posizione percepita viene processata dal cervello attraverso le cosiddette differenze interaurali. In pratica, in base a come il suono arriva all'ascoltatore, le due orecchie ricevono due segnali diversi l'uno dall'altro. Le differenze rilevanti sono in termini di livello (ILD – Interaural Level Difference), di tempo (ITD – Interaural Time Difference) e di fase (IPD – Interaural Phase Difference). Queste differenze sono praticamente nulle sul piano mediano, ne consegue una scarsa abilità dell'uomo a riconoscere la provenienza dei suoni se posizionati proprio su tale piano. A questo proposito Blauert [17] ha dimostrato come la percezione della posizione di toni puri sia dipendente quasi esclusivamente dalla frequenza: alcuni toni vengono percepiti nella parte frontale, altri nel posteriore e alcuni sopra la testa anche se emessi da altoparlanti con direzione diversa, ma sempre sul piano mediano.

1.3.1 | Minimum Audible Angle (MAA)

Per quantificare la performance uditiva di localizzazione si usa abitualmente il minimum audible angle (MAA) definito da Mills [18] come “la più piccola differenza percepibile tra gli azimut di due sorgenti identiche”, concetto facilmente estendibile all'angolo di elevazione. Lo studio di Grantham et al. [19], mostra valori di MAA in azimut, elevazione e diagonale (60° rispetto all'orizzontale) riportati in Tab. 1. I dati in questione evidenziano come l'uomo riesca a distinguere molto meglio due sorgenti singole quando si differenziano angolarmente in azimut, risultati che confermano quelli trovati da Perrot e Saberi [20] in precedenza; in tutti i casi è stato preso il riferimento, ovvero la posizione attorno alla quale si

Tab. 1 – Valori di MAA riportati in di Grantham et al. [19]; Wideband=70-15000 Hz, High pass=6000-15000Hz, Low pass=70-2000 Hz; fra parentesi la deviazione standard MAA values from Grantham et al. [19]; Wideband=70-15000 Hz, High pass=6000-15000Hz, Low pass=70-2000 Hz; standard deviation in brackets

Segnale	Azimut	Diagonale	Verticale
Wideband	1.6° (0,3°)	2.8° (0,4°)	6.5° (2.4°)
High pass	1.6° (0,6°)	4.1° (1.4°)	11.6° (12.0°)
Low pass	1.5° (0,4°)	3.4° (1.0°)	19.0° (13.0°)

presenteranno stimoli di confronto appena spostati per ricavare i valori di MAA, a 0° di elevazione ed azimut, che successivamente verrà abbreviato in “direzione frontale”. Altri studi, come quello di Mills [18] e di Perrot e Saberi [20], rilevano un MAA nella direzione frontale minore rispetto a 1.6° di wide-band-azimut, ovvero di circa 1° in azimut.

Tuttavia, gli MAA ricavati con sistemi che prevedono sorgenti virtuali, come Ambisonics, sono tendenzialmente più alti rispetto a quelli citati in precedenza che sono stati trovati con l'utilizzo esclusivo di sorgenti reali. Nello studio di Sochaczewska et al. [21] si vede come, con un Ambisonics del terzo ordine, il valore MAA nella direzione frontale si aggiri attorno ai 3.2° di media. In ogni caso viene riportato un andamento di MAA al variare del riferimento in azimut molto simile: MAA cresce leggermente fino ai 60°-75° di azimut di riferimento per poi aumentare in maniera più decisa fino al riferimento 90° in cui si raggiungono, per lo studio [21], valori di MAA maggiori di 10° in media.

Invece, Meng et al. [22] hanno raccolto valori di MAA utilizzando VBAP. Lo studio dimostra come si possa sostituire la sorgente fisica con quella virtuale sul piano orizzontale nelle posizioni azimutali di 0° e 90°. Infatti, le misure medie di MAA con sorgente virtuale sono di 1.1° a 0° e di 3.1° a 90°, mentre quelle con sorgenti fisiche sono rispettivamente di 1.2° e 3.3°.

In conclusione, occorre tenere conto dell'influenza sulla percezione sonora di un dispositivo “indossato” in testa, come il visore per realtà virtuale, e l'effetto dell'informazione visiva fornita. È stato dimostrato da Huisman et al. [8] che, sul piano orizzontale:

- La prestazione di localizzazione non presenta differenze significative tra terzo, quinto e undicesimo ordine Ambisonics; peggiora molto per il primo.
- In assenza di informazione visiva, la presenza del visore sposta la percezione della sorgente più a sinistra quando essa è a sinistra e, rispettivamente, a destra.
- In presenza dell'informazione visiva, l'impatto del visore è trascurabile sulla capacità di localizzazione per applicazioni “ludiche”, non compromette l'esperienza immersiva, ma è da tenere in conto per la sperimentazione scientifica.

2 | Realizzazione della “Camera dei Diamanti”

2.1 | Hardware

Nel laboratorio di acustica dell'Università di Ferrara era presente una cabina audiometrica di dimensioni esterne di 4.20 × 3.18 × 2.76 m, prodotta da Puma Acoustics e trattata acusticamente con coni di melammina. Il pavimento della stanza è sollevato da terra e sostenuto da supporti che permettono un isolamento vibrazionale dai disturbi che possono incorrere dall'ambiente esterno. Inoltre, è anche presente un sistema di ventilazione estremamente silenzioso che consente un prezioso e fondamentale ricircolo di aria. Per disporre gli altoparlanti su di una superficie quasi sferica, è stato montato, in prima battuta, un telaio in acciaio che avesse la funzione di sostegno di pianta quadrata di lato 3.26 m e alto 2.20 m (vedi Fig. 4).

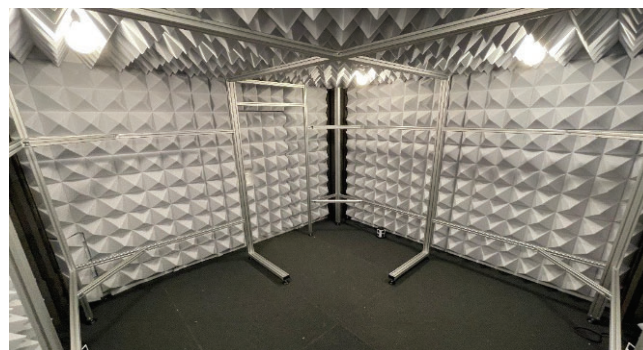


Fig. 4 – Cabina audiometrica con telaio in acciaio
Audiometric booth with steel frame

Successivamente, sono stati fissati sul telaio 32 altoparlanti passivi e 8 altoparlanti attivi sono stati posti su supporti appositi. La scelta del diverso montaggio degli altoparlanti è dovuta principalmente alla necessità di sfruttare e riempire lo spazio lontano dalla struttura. In questo modo si sono anche prevenuti possibili problemi di fissaggio degli altoparlanti attivi che sono molto pesanti (vedi Tab. 2). Inoltre, sapendo che le casse passive in dotazione coprono un range in frequenza minore rispetto alle attive, questa sistemazione ha permesso di posizionare gli altoparlanti più performanti nella parte frontale sul piano orizzontale, ovvero la zona in cui è richiesta la maggiore risoluzione angolare per via dei valori di MAA più bassi.

Il passo successivo è stato fissare al telaio un ulteriore strato di prismi di melammina per cercare di abbassare la frequenza di taglio della camera e quindi prevenire la maggior parte delle riflessioni; vedi Fig. 5.

Inoltre, la camera è stata dotata di una sedia regolabile in altezza, posta al centro della pianta, per fare in modo che l'ascoltatore possa posizionarsi con le orecchie a 1,12 m da terra, punto corrispondente al centro dell'array di casse, per avere la migliore esperienza di riproduzione possibile.



Fig. 5 – Camera dei Diamanti” ultimata
Finished “Camera dei Diamanti”

In Tab. 2 sono riportate le specifiche degli altoparlanti utilizzati e si vuol far notare come la riproduzione copra un range di frequenza vasto. Il subwoofer è stato aggiunto in seguito per “aiutare” le basse frequenze.

Tab. 2 – Specifiche degli altoparlanti
Loudspeakers' specs

	Tannoy 7 Gold (active)	Tannoy Precision 6D (active)	RCF Compact M04 (passive)
Potenza max	300 W	110 W	240 W
SPL max	110 dB	116 dB	113 dB
Range di frequenza	65 Hz 20 kHz	59 Hz 51 kHz	90 Hz 20 kHz
Frequenza di crossover	1500 Hz	2500 Hz	1800 Hz
Tipo di crossover	Two-way	Two-way	Two-way
Peso	9.4 kg	13 kg	2.3 kg

In Fig. 6 è riportata la disposizione degli altoparlanti nella camera. I cerchi si riferiscono alla posizione degli altoparlanti passivi, mentre le piccole stelle a quelle degli attivi.

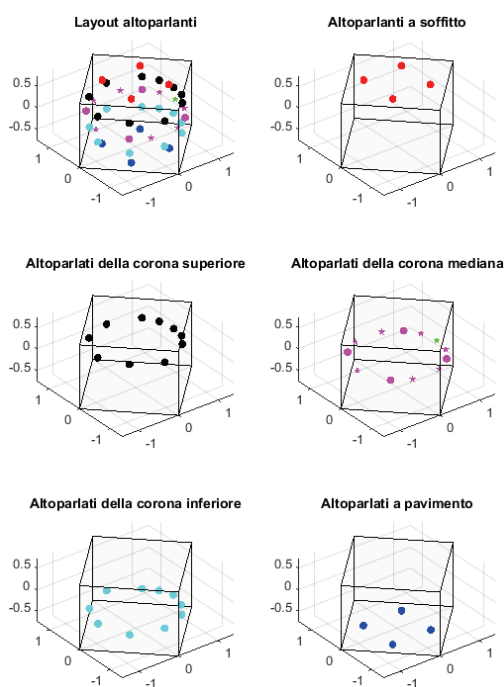


Fig. 6 – Disposizione degli altoparlanti
Loudspeakers' layout

Si evidenzia come la distribuzione degli altoparlanti non sia omogenea, ma abbia una densità più elevata nella parte frontale, ovvero nel punto indicato con la stella verde, coincidente sia

con un altoparlante attivo che con il punto di origine (0,0) del sistema di riferimento azimuth-elevazione riferito alla camera. Una maggiore densità permette una migliore capacità del sistema a spazializzare il suono in quella zona, che bisogna ricordare essere la più critica da trattare dato che ivi il valore di MAA è il minimo.

Il cablaggio fisico è rappresentato dalle frecce che uniscono i vari blocchi in Fig. 7. Quindi, il segnale parte da un computer, passa attraverso una scheda audio RME Fireface UFX III e si separa in due filoni che si dirigono verso gli altoparlanti attivi e passivi. In quest'ultimo, il segnale viene trattato da un amplificatore digitale a 32 canali Innosonix MA32/LP².

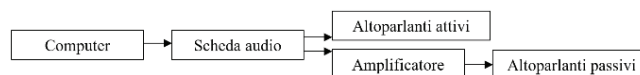


Fig. 7 – Cablaggio tra componenti hardware
Hardware components wiring

2.2 | Software

Il controllo dei segnali audio è stato sviluppato con Plogue Bidule, un software DAW (Digital Audio Workstation) nel quale è stata implementata una catena di “plugins” che permettesse la restituzione acustica desiderata all'interno della camera. I plugins sono estensioni del programma, spesso create da terze parti, che aggiungono funzioni non presenti nel software in sé. Possono essere intesi come una sorta di filtro sul segnale di input, l'output sarà modificato secondo alcuni parametri. In questo studio verranno utilizzate le suite COMPASS e SPARTA sviluppate dall'Università di Aalto, la IEM Plugin Suite dell'Università di Graz e il plugin MCFX-Convolver di Matthias Kronlachner e poi ottimizzato da Angelo Farina.

2.2.1 | Flusso dei dati e filtri

In Fig. 8 è riportato un esempio del flusso dei dati che parte da un file audio (.wav da 64 canali) di input e che arriva all'output della scheda audio.

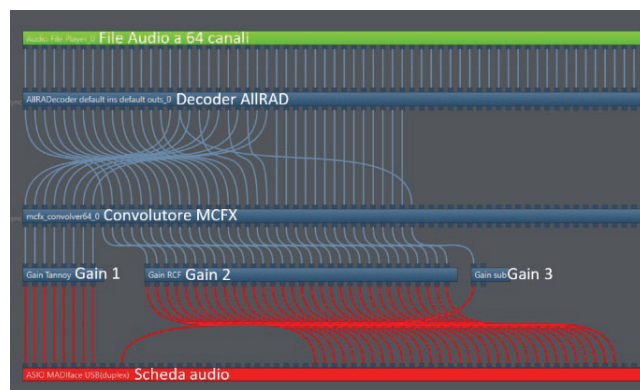


Fig. 8 – Flusso dei dati in Plogue Bidule, dove Gain 1, 2 e 3 si riferiscono rispettivamente ai gain di Tannoy, RCF e subwoofer
Data flow in Plogue Bidule where Gain 1, 2 and 3 refer respectively to gains of Tannoy, RCF and subwoofer

Il segnale passa, inizialmente, attraverso il decoder AI-IRADecoder (IEM Plugins Suite), ovvero l'artefice della distribuzione del segnale sui vari altoparlanti sfruttando l'algoritmo AllRAD citato in precedenza. Poi, il segnale viene convoluto con filtri per avere l'equalizzazione e l'allineamento temporale ad un campione rispetto al centro della camera di ciascuna cassa. Il segnale è campionato a 48 kHz, ciò significa che il tempo massimo di differenza con cui due segnali possono arrivare al centro della camera è di circa 20 μ s. In questo modo le casse emettono come se fossero posizionate su di una sfera quasi perfetta. Successivamente il segnale entra in blocchi che gestiscono i livelli di ciascuna cassa i quali sono regolati in modo che ogni altoparlante produca lo stesso livello misurato al centro della stanza.

Sia i filtri che i livelli sono stati misurati e realizzati con una capsula microfonica a condensatore Brüel & Kjær (B&K) 4165 associata ad un preamplificatore B&K 5935. Gli "intrecchi" dei collegamenti in Fig. 8 sono dovuti alla diversa numerazione tra filtri e canali degli altoparlanti.

3 | Verifica delle prestazioni di spazialità del suono

3.1 | Metodologia

Le performance oggettive della camera, in termini di rappresentazione spaziale del suono, sono state studiate attraverso la riproduzione di un rumore da una sorgente singola e la registrazione di esso attraverso un array microfonico. Confrontando la posizione della sorgente del rumore in input con quella rilevata dal microfono, si può stimare l'errore commesso.

Sono state verificate le posizioni di sorgenti singole sia virtuali che reali. Per le prime sono stati sfruttati dei file.wav di rumori rosa in terzi d'ottava (bande con frequenza di centro banda di 400 Hz, 1000 Hz e 4000 Hz) e uno a banda larga resi disponibili dall'Università di York (Cooper [23]) e codificati con un Ambisonics del settimo ordine. Invece, la verifica del posizionamento della sorgente reale è stata effettuata utilizzando un rumore bianco (100 Hz-15 kHz) emesso da un altoparlante alla volta.

Posizioni Sorgenti Virtuali	
Azimut	Elevazione
0	2
0	90
0	-90
-90	2
90	2
-180	0
0	26
-36	-26
36	-26
-72	26
72	26
-108	-26
108	-26
-144	26
144	26
180	-26

Fig. 9 – Posizioni delle sorgenti virtuali (Cooper [23]) espresse in gradi
Virtual sources positions (Cooper [23]) expressed in degrees

Le coordinate delle sorgenti virtuali (espresse in gradi, vedi Fig. 9) coincidono con le facce di un dodecaedro e di un cubo, non sono 18 (12+6), ma 16 perché due di esse coincidono. L'elevazione di qualche sorgente è a 2° e non a 0° (ad esempio la faccia frontale ce la si aspetterebbe a 0° sia di azimut che di elevazione) per un probabile errore nella codifica del segnale, ma le misure non risentono di questo scostamento.

3.2 | Misura

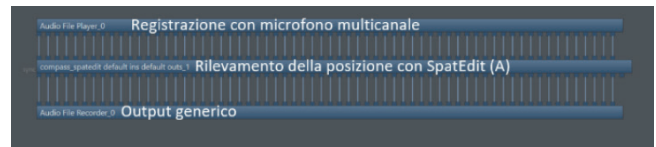


Fig. 10 – Catena di plugins per la misura della posizione della sorgente singola registrata con Eigenmike® a 64 canali
Plugins chain for measuring the position of a single source recorded with a 64-channel Eigenmike®

In Fig. 10 è mostrata la catena di plugin utilizzata per ricavare la posizione della sorgente singola dalla registrazione effettuata con un microfono Eigenmike® a 64 canali mostrato in Fig. 12. In input è presente la registrazione che viene analizzata dal plugin SpatEdit (A) della suite COMPASS in Fig. 11.

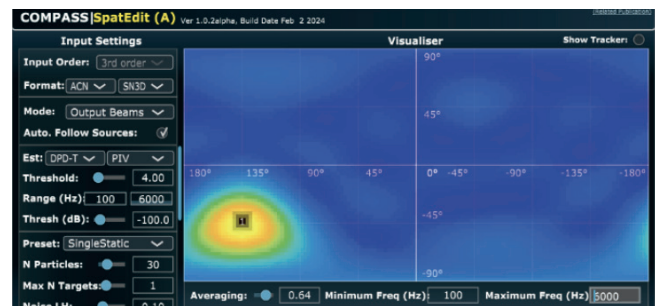


Fig. 11 – COMPASS SpatEdit (A)
COMPASS SpatEdit (A)

Il componente in questione prende il segnale in ingresso e calcola una colormap dell'ambiente acustico. È fornito di un tracker che può seguire automaticamente la posizione delle sorgenti presenti nella scena sonora e fornire una stima delle coordinate in azimut ed elevazione.

È stato utilizzato il preset "SingleStatic" che ottimizza i valori dei parametri che governano il tracking per una sorgente singola stazionaria. Il tracking automatico è stato sfruttato fintanto che non commettesse errori molto evidenti: se il quadrato (con al suo interno il numero 1, vedi Fig. 11) che stima la posizione della sorgente era molto spostato rispetto "all'alone" più chiaro, si è preferito il metodo manuale trascinando il quadrato stesso al centro della zona più gialla.

L'errore è stato calcolato come la differenza tra le posizioni registrate e quelle emesse in termini di distanza lineare.

Questa non è altro che una corda della sfera e quindi è stato possibile riformulare l'errore calcolato come angolo al centro.



Fig. 12 – Setup di misura con Eigenmike® a 64 canali
Measure setup with 64 channels Eigenmike®

3.3 | Risultati

I risultati sono stati calcolati per tutte le posizioni di sorgente singola virtuale e reale come spiegato in precedenza. Di seguito sono riportate le colormaps ottenute con un'interpolazione di tipo “natural neighbor” dell'errore di ogni punto valutato come angolo al centro. Inoltre, i grafici mostrano con cerchi le posizioni reali delle sorgenti e con crocette le posizioni registrate. In Fig. 13 sono riportati gli assi di azimuth ed elevazione per facilitare la lettura dei grafici in Fig. 15 e Fig. 14.

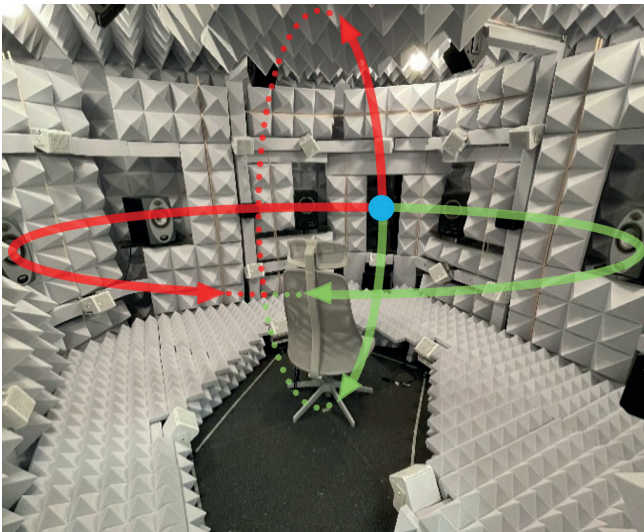


Fig. 13 – Rappresentazione degli assi in riferimento al sistema azimuth-elevazione riportato nei grafici a seguire
Representation of the azimuth-elevation axes used in the following graphs

Dal grafico in Fig. 15 si può evincere come gli errori siano sempre tra 1° e 3°. Inoltre, si vuole evidenziare come nella zona critica, ovvero nell'intersezione degli assi, l'errore sia limitato a 1.3°.

Ancora una volta, facendo riferimento alla Fig. 14, l'errore non supera i 3°, risulta mediamente più elevato e anche distribuito in maniera più omogenea sulla parte di superficie sferica interpolata. È da notare anche un'assenza di evidenti pattern di errore se non per un accenno di somiglianza nell'emisfero destro (Fig. 14 (b), (c) e (d)).

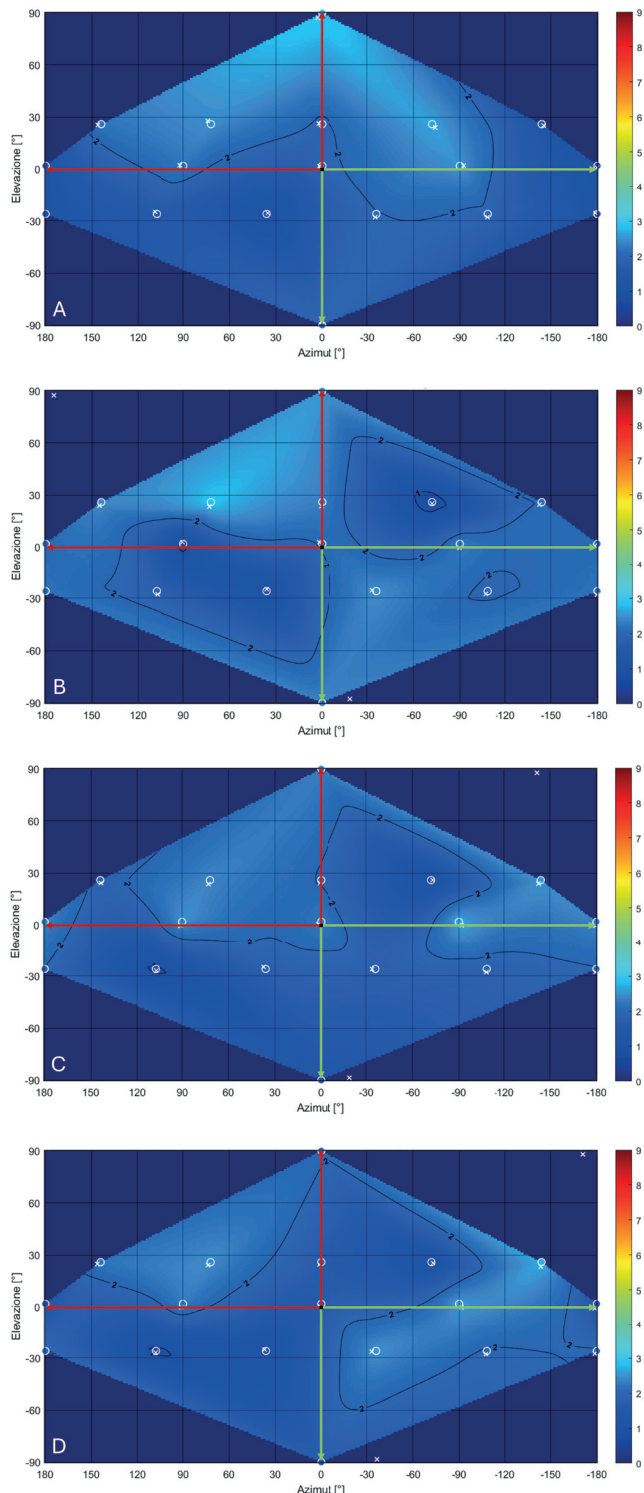


Fig. 14 – Colormaps dell'errore delle sorgenti virtuali nel caso di rumore rosa: per bande di terzi d'ottava (frequenze di centro banda di 250 Hz (a), 1000 Hz (b) e 4000Hz (c)) e per banda larga (d)
Error colormaps of virtual sources for different pink noise types: one-third-octave bands (center-band frequency equals to 250 Hz (a), 1000 Hz (b) e 4000Hz (c)) and wide band (d)

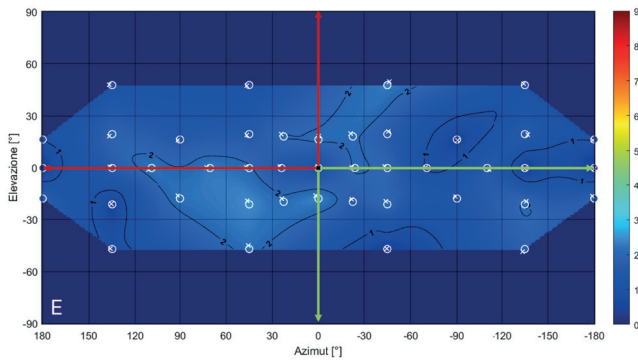


Fig. 15 – Colormaps dell'errore delle sorgenti reali con rumore bianco
Error colormaps of real sources using a white noise

Nel grafico in Fig. 16 si riportano le medie degli errori di-
visi per tipologia di segnale, di sorgente e di emisfero a cui
appartiene la sorgente singola.

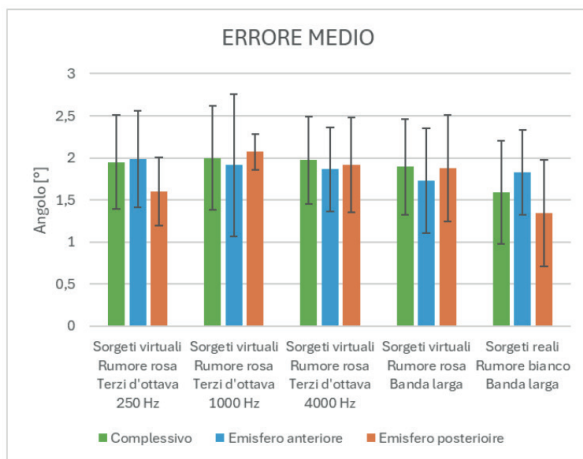


Fig. 16 – Confronto tra diversi gruppi di errore divisi per tipologia di segnale e di sorgente con sottogruppi relativi all'emisfero in cui sono posizionate le sorgenti singole
Comparison between errors grouped by signal and source type with subgroups indicating in which hemisphere the sound comes from

3.4 | Discussione

I risultati ottenuti consentono di poter affermare che la Camera dei Diamanti possiede ottime capacità di riprodurre uno stimolo sonoro localizzato. Infatti, i valori di errore angolare, misurato “oggettivamente”, sono comparabili con i valori di MAA di letteratura. Poi confrontando i valori medi degli errori divisi per tipologia, si può vedere come non siano presenti grosse differenze tra valori ottenuti con segnali diversi, emisfero di appartenenza della sorgente che sia essa reale o virtuale.

Nella valutazione dell'errore non sono state considerate le incertezze e i limiti che una misura del genere può introdurre nella stima della posizione di una sorgente sonora. Di conseguenza la misura è effettivamente comprensiva di una parte di errore che non è stata stimata e potrebbe incidere in termini migliorativi.

Inoltre, è necessario tener presente che le performance stimate sono prettamente oggettive. Questo non assicura una performance percettiva ottimale, ma garantisce il buon funzionamento dell'hardware. È una validazione di base per poter verificare che ci siano condizioni che possano consentire un ascolto spazializzato soddisfacente.

4 | OSC e sviluppo del sistema a realtà virtuale

4.1 | OSC e WhisPER

È stata sviluppata una comunicazione OSC (Open-Sound-Control, sviluppato da Wright e Freed [24]) per lo scambio di informazioni tra player DAW e altri software. Il protocollo di comunicazione scelto consente di avere un'ampia gamma di controlli che rendono il sistema adattabile alle più svariate applicazioni.

OSC è stato utilizzato per far mandare comandi al DAW da Unity, software che gestisce la parte visiva della realtà virtuale, oppure da WhisPER [25], un applicativo MATLAB sviluppato dall'Università Tecnica di Berlino che offre la possibilità di eseguire numerosi test psicoacustici di diversa natura. In Fig. 17 è mostrato un test per la valutazione di soglie psicofisiche. Per esempio, si potrebbe chiedere al soggetto di riconoscere il suono percepito come “diverso” rispetto ad altri due uguali. In particolare, in Fig. 17, è illustrato un test a 3-AFC (risposta forzata tra tre alternative) con procedura adattiva 2down/1up [26], ovvero lo stimolo corretto da scegliere diventa più “difficile” da distinguere quando viene scelto correttamente per due volte di fila e torna ad essere più “facile” non appena viene selezionata la risposta sbagliata.

Questa procedura può essere, per esempio, implementata per la valutazione del minimum audible angle (MAA).

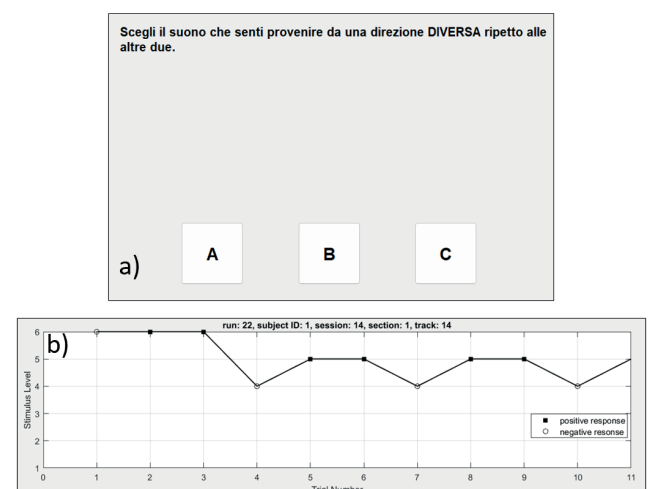


Fig. 17 – Esempio di schermata WhisPER per un test a 3AFC (a) per stabilire una soglia uditiva con procedura 2down/1up (b)
Example of a WhisPER screenshot for a 3AFC test (a) to stabilize a hearing threshold with the 2down/1up procedure (b)

Il sistema a realtà virtuale utilizza un visore Meta Quest 3 per cui sono state sviluppate tre demo in ambiente Unity in collaborazione con l'Ufficio di Servizi E-learning e Multimediali dell'Università di Ferrara.

- Esperienza audio-visiva virtuale (senza interazione).
Vengono proposti un video bidimensionale “proiettato” su una sfera tridimensionale nera e la relativa traccia audio con Ambisonics del settimo ordine. L'esperienza introduce il partecipante al mondo dell'audio spazializzato sia dal punto di vista percettivo che concettuale in quanto l'avatar presente in Fig. 18 spiega l'evoluzione di questa tecnologia sfruttando l'Ambisonics stesso.

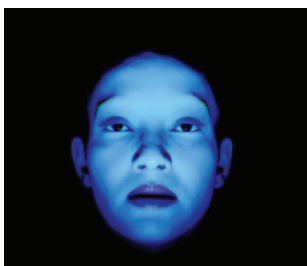


Fig. 18 – Speech 70A (Cooper et al. [23])
Speech 70A (Cooper et al. [23])

- Test di localizzazione con puntatore.
Il test in questione permette di stimare le differenze tra la posizione della sorgente sonora singola e quella percepita dall'ascoltatore, valori utili alla ricostruzione delle capacità di localizzazione del soggetto.
Il soggetto è in grado di comunicare al sistema la direzione percepita grazie ad un puntatore virtuale che funziona come una “pistola” (vedi Fig. 19). In pratica viene definita la posizione di una sorgente singola (da cui verrà emesso lo stimolo sonoro) con il plugin SPARTA AmbiENC attraverso comandi OSC inviati da un programma terzo. Il partecipante allora deve sparare nella direzione nella quale sente arrivare il suono; a quel punto la sorgente cambia posizione randomicamente tra un set prefissato di punti e il soggetto tornerà a sparare. Ad ogni iterazione viene calcolata la differenza tra la posizione della sorgente e quella indicata dal soggetto.

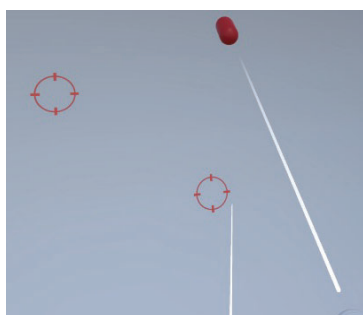


Fig. 19 – Test di localizzazione
Localization test

- Test di intelligibilità della parola a matrice contestualizzato.
Il test a matrice è uno strumento clinico per la valutazione dell'intelligibilità della parola, inizialmente ideato da Hagerman [27] e poi riproposto in lingua italiana da Puglisi et al. [28]. Al soggetto vengono presentati un rumore “disturbante” e frasi di cinque parole pronunciate da un parlatore. Il partecipante deve allora selezionare sulla matrice le parole percepite.

Per la demo è stata registrata una scena di disturbo al bar della facoltà di ingegneria con una videocamera a 360° e un array microfonico Zylia®. Poi, è stato raccolto il materiale audio-video del parlatore in camera anecoica. La figura è stata sovrapposta al video 3D e la traccia audio anecoica è riprodotta in parallelo alla scena acustica spazializzata. Alla fine del test viene prodotto automaticamente file di report che segnala le risposte sbagliate e i tempi di risposta del soggetto.



Fig. 20 – Test a matrice di intelligibilità della parola contestualizzato
Contextualized speech intelligibility matrix test

5 | Prospettive future

Sono in programma miglioramenti del sistema, ovvero l'inserimento di un sistema che consenta di avere pupillometria e una routine di calibrazione del visore per la realtà virtuale. Dal comportamento della pupilla si può stimare l'affaticamento di un soggetto nel portare a termine un compito [29]. Dati del genere sono fondamentali nella valutazione della differente “fatica uditiva” che può incorrere in diversi pazienti con stesse capacità di intelligibilità della parola [30]. La calibrazione, invece, si concretizza nel far corrispondere il sistema di riferimento virtuale con quello reale, quindi quello del visore con quello mostrato in Fig. 13. Una collimazione precisa tra i due

sistemi di riferimento garantisce, ovviamente, raccolte dati più accurate, ad esempio quando si svolgono test con puntatore del tipo in Fig. 19.

La “Camera dei diamanti” è uno strumento che può essere sfruttato in una vasta gamma di applicazioni. Si prospettano ricerche in campo medico-psicologico, come la valutazione della percezione della localizzazione di pazienti con protesi cocleari in ottica di insegnamento e riabilitazione.

Invece, un esempio di impiego a livello automotive è il perfezionamento dei sistemi AVAS (Acoustic Vehicle Alerting Systems): ricreando uno scenario acustico cittadino si possono valutare le migliori strategie di progettazione del suono tenendo conto del contesto e della direzione di arrivo del pericolo con grande realismo percettivo. Inoltre, si possono ricreare le condizioni all'interno dell'abitacolo per simulare il rumore percepito a diversi regimi di marcia e progettare soluzioni mirate al miglioramento dell'ambiente acustico.

Le applicazioni industriali comprendono la sound quality del prodotto, in particolare la cura dell'emissione sonora dell'oggetto visto da direzioni diverse, ma anche in stanze con caratteristiche acustiche particolari (ad esempio molto riverberanti).

Un altro ambito interessante è l'utilizzo di questo sistema in campo e-learning, sia dal punto di vista didattico-scolastico che da quello di training di sicurezza. La realtà virtuale ad alta fedeltà può essere una componente decisiva nel preparare le persone alla gestione di scenari critici. In tal modo sarebbe possibile ricreare una situazione molto realistica in cui il soggetto riesce a provare le procedure di sicurezza senza un vero pericolo.

6 | Conclusioni

Lo sviluppo di una metodologia di validazione oggettiva delle prestazioni di spazializzazione è cruciale per il confronto tra performance di camere differenti, per evidenziare carenze (come valutare eventuali riflessioni e altoparlanti mal orientati), ma anche per rendersi conto dei limiti di risoluzione spaziale del sistema.

La “Camera dei Diamanti” è una struttura innovativa in grado di porsi in una posizione strategica nello spazio della ricerca di base e applicata, ma anche dell'industria. Le potenzialità multipurpose garantiscono la possibilità di intraprendere plurimi percorsi che portino innovazione a realtà medicali e d'impresa. Inoltre, questa tecnologia è scalabile e adattabile alla precisa applicazione: si può ad esempio pensare di ricostruire un sistema del genere attorno ad un simulatore NVH (Noise, Vibration and Harshness) automotive o anche migliorare e aggiornare le cabine audiologiche presenti nelle cliniche e negli ospedali.

Conclusions

The introduction of a new technique capable of objectively validating the sound spatialization performance of the chamber is crucial to compare other rooms, to highlight localized deficiencies (for exam-

ple, to evaluate reflections and disoriented loudspeakers), and be aware of the internal limitations of the audio system itself.

The “Camera dei Diamanti” is a new, innovative, and powerful tool that occupies a strategic position for basic and applied research, but also in the industrial world. Its multipurpose functions guarantee the possibility of exploring many different avenues of innovation for human well-being and technical progress. Moreover, this technology is scalable and adaptable to specific applications: it is possible, for example, to build the hardware all around an NVH (Noise, Vibration and Harshness) vehicle simulator or to improve the performance of audiometric booths already in use in clinics and hospitals.

7 | Ringraziamenti

Si ringraziano il gruppo di ricerca di acustica dell'Università di Ferrara, l'Ufficio di Servizi E-learning e Multimediali dell'Università di Ferrara, in particolare Andrea Bertelli e Giulia Tampone che hanno sviluppato le applicazioni per il visore per la realtà virtuale, Angelo Farina, il Comitato Editoriale della RIA e il team di FrancoAngeli Editore.

Il presente lavoro è stato sostenuto dal fondo FIRD2024 “Camera dei Diamanti: una piattaforma UniFe di testing e training in realtà virtuale”.

Bibliografia

- [1] N.A. Lesica, Why Do Hearing Aids Fail to Restore Normal Auditory Perception?, *Trends in Neurosciences* 41 (2018) 174–185. <https://doi.org/10.1016/j.tins.2018.01.008>.
- [2] V. Hohmann, R. Paluch, M. Krueger, M. Meis, G. Grimm, The Virtual Reality Lab: Realization and Application of Virtual Sound Environments, *Ear & Hearing* 41 (2020) 31S–38S. <https://doi.org/10.1097/AUD.0000000000000945>.
- [3] Y.-H. Wu, E. Stangl, O. Chipara, S.S. Hasan, S. DeVries, J. Oleson, Efficacy and Effectiveness of Advanced Hearing Aid Directional and Noise Reduction Technologies for Older Adults With Mild to Moderate Hearing Loss, *Ear & Hearing* 40 (2019) 805–822. <https://doi.org/10.1097/AUD.0000000000000672>.
- [4] R.A. Bentler, Effectiveness of Directional Microphones and Noise Reduction Schemes in Hearing Aids: A Systematic Review of the Evidence, *J Am Acad Audiol* 16 (2005) 473–484. <https://doi.org/10.3766/jaaa.16.7.7>.
- [5] M.T. Cord, R.K. Surr, B.E. Walden, O. Dyrland, Relationship between Laboratory Measures of Directional Advantage and Everyday Success with Directional Microphone Hearing Aids, *J Am Acad Audiol* 15 (2004) 353–364. <https://doi.org/10.3766/jaaa.15.5.3>.
- [6] G. Llorach Tó, G. Grimm, M. Hendrikse, V. Hohmann, Towards Realistic Immersive Audiovisual Simulations for Hearing Research: Capture, Virtual Scenes and Reproduction, 2018. <https://doi.org/10.1145/3264869.3264874>.
- [7] G. Grimm, J. Lubradzka, V. Hohmann, A Toolbox for Rendering Virtual Acoustic Environments in the Context of Audiology, *Acta Acustica United with Acustica* 105 (2019) 566–578. <https://doi.org/10.3813/AAA.919337>.
- [8] T. Huisman, A. Ahrens, E. MacDonald, Ambisonics Sound Source Localization With Varying Amount of Visual Information

- in Virtual Reality, *Frontiers in Virtual Reality* 2 (2021). www.frontiersin.org/journals/virtual-reality/articles/10.3389/frvir.2021.722321.
- [9] A. Guastamacchia, R.G. Rosso, G.E. Puglisi, F. Riente, L. Shtrepi, A. Astolfi, Real and Virtual Lecture Rooms: Validation of a Virtual Reality System for the Perceptual Assessment of Room Acoustical Quality, *Acoustics* 6 (2024) 933–965. <https://doi.org/10.3390/acoustics6040052>.
- [10] F. Pausch, G. Behler, J. Fels, SCaLAR – A surrounding spherical cap loudspeaker array for flexible generation and evaluation of virtual acoustic environments, *Acta Acust.* 4 (2020) 19. <https://doi.org/10.1051/aacus/2020014>.
- [11] G.D. Romigh, D.S. Brungart, B.D. Simpson, Free-Field Localization Performance With a Head-Track Virtual Auditory Display, *IEEE J. Sel. Top. Signal Process.* 9 (2015) 943–954. <https://doi.org/10.1109/JSTSP.2015.2421874>.
- [12] F. Zotter, M. Frank, Ambisonic Amplitude Panning and Decoding in Higher Orders, in: F. Zotter, M. Frank (Eds.), *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, Springer International Publishing, Cham, 2019: pp. 53–98. https://doi.org/10.1007/978-3-030-17207-7_4.
- [13] F. Zotter, M. Frank, All-Round Ambisonic Panning and Decoding, *Journal of the Audio Engineering Society* 60 (2012) 807–820.
- [14] V. Pulkki, *Spatial Sound Generation and Perception by Amplitude Panning Techniques*, (2001).
- [15] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, The MIT Press, 1996. <https://doi.org/10.7551/mitpress/6391.001.0001>.
- [16] A. Carlini, C. Bordeau, M. Ambard, Auditory localization: a comprehensive practical review, *Frontiers in Psychology* 15 (2024). www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2024.1408073.
- [17] J. Blauert, Sound Localization in the Median Plane, *Acta Acustica United with Acustica* 22 (1969).
- [18] A.W. Mills, On the Minimum Audible Angle, *The Journal of the Acoustical Society of America* 30 (1958) 237–246. <https://doi.org/10.1121/1.1909553>.
- [19] W. Grantham, B. Hornsby, E. Erpenbeck, Auditory spatial resolution in horizontal, vertical, and diagonal planes, *The Journal of the Acoustical Society of America* 114 (2003) 1009–22. <https://doi.org/10.1121/1.1590970>.
- [20] D.R. Perrott, K. Saberi, Minimum audible angle thresholds for sources varying in both elevation and azimuth, *The Journal of the Acoustical Society of America* 87 (1990) 1728–1731. <https://doi.org/10.1121/1.399421>.
- [21] K. Sochaczewska, P. Malecki, M. Piotrowska, Evaluation of the Minimum Audible Angle on Horizontal Plane in 3rd order Ambisonic Spherical Playback System, 2021. <https://doi.org/10.1109/I3DA48870.2021.9610858>.
- [22] R. Meng, J. Xiang, J. Sang, C. Zheng, X. Li, S. Bleeck, J. Cai, J. Wang, Investigation of an MAA Test With Virtual Sound Synthesis, *Frontiers in Psychology* 12 (2021). www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.656052.
- [23] J. Cooper, Immersive Audiovisual Materials Database, (2024). <https://doi.org/10.5281/ZENODO.10571315>.
- [24] M. Wright, A. Freed, OSC, (2021). <https://ccrma.stanford.edu/groups/osc/index.html> (accessed December 5, 2024).
- [25] S. Ciba, A. Wlodarski, H.-J. Maempel, WhisPER – A New Tool for Performing Listening Tests, 126th Audio Engineering Society Convention 2009 1 (2012).
- [26] H. Levitt, Transformed Up-Down Methods in Psychoacoustics, *The Journal of the Acoustical Society of America* 49 (1971) Suppl 2:467+. <https://doi.org/10.1121/1.1912375>.
- [27] B. Hagerman, Sentences for Testing Speech Intelligibility in Noise, *Scandinavian Audiology* 11 (1982) 79–87. <https://doi.org/10.3109/01050398209076203>.
- [28] G.E. Puglisi, A. Warzybok, S. Hochmuth, C. Visentin, A. Astolfi, N. Prodi, B. Kollmeier, An Italian matrix sentence test for the evaluation of speech intelligibility in noise, *International Journal of Audiology* 54 (2015) 44–50. <https://doi.org/10.3109/14992027.2015.1061709>.
- [29] J. Beatty, Task-evoked pupillary responses, processing load, and the structure of processing resources., *Psychological Bulletin* 91 (1982) 276–292. <https://doi.org/10.1037/0033-2909.91.2.276>.
- [30] M.B. Winn, D. Wendt, T. Koelewijn, S.E. Kuchinsky, Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started, *Trends in Hearing* 22 (2018) 2331216518800869. <https://doi.org/10.1177/2331216518800869>.

