**Slide**

Il nostro progetto consiste nel lavorare con una rete neurale applicata all’audio 3d.

In particolare, nel nostro lavoro abbiamo usato una rete nel dominio quaternionico.

Dato in input un frammento di file audio, il task che la rete deve risolvere è quello di saper classificare la sorgente audio (capire a quale classe appartiene, per esempio “cane che abbaia”, “porta che sbatte”, “risata” e così via..) e di localizzarlo (quind di capire la direzione di arrivo e la distanza dalla sorgente in uno spazio tridimensionale).

**Slide**

Sappiamo che in ambienti aperti o in stanze senza eco, è facile stimare la direzione di arrivo di un suono attraverso il vettore dell’intensità attiva. Prendendo la direzione opposta a quella di questo vettore, infatti, si ricava proprio la doa.

[*La matematica dietro tutti questi termini verrà spiegata nel dettaglio tra qualche slide*]

Quello che vogliamo evidenziare qui è che, nel momento in cui disponiamo di una formula esatta per il calcolo della direzione di arrivo, perché si ha bisogno di una rete neurale?

**Slide**

Nei casi pratici è molto frequente di trovarsi in presenza di ambienti rumorosi o con riverbero.

In questi casi il suono potrebbe aver preso molti percorsi imprevedibili per andare dalla sorgente all’ascoltatore.

Questo porta il vettore dell’intensità attiva ad essere inconsistente e non può essere usato per trarre una stima precisa della doa.

È stato provato, invece, che in questi casi, passando in input le features appropriate, approcci basati su reti neurali offrono risultati migliori.

**Slide**

I file audio che abbiamo usato sono stati registrati con un microfono chiamato “first-order ambisonics”.

Questo microfono ha la particolarità di essere composto da 4 capsule coincidenti e riesce a scomporre il suono in una base ortogonale di armoniche sferiche.

Delle 4 capsule la prima è detta di ordine 0, (quella indicata con W), è omnidirezionale ed ha guadagno unitario in ogni direzione.

Le altre 3 capsule sono dette di “primo ordine”, sono quelle che nella foto sono a forma di 8, sono disposte in modo ortogonale tra di loro e sono indicate con X, Y, Z.

**Slide**

Queste quattro componenti possono essere viste come le componenti di un Quaternione di Hamilton.

Un Quaternione è un numero iper-complesso formato da una componente reale e tre componenti immaginarie.

In questo caso possiamo porre la componente W come parte reale, e tre componenti X,Y,Z come le componenti immaginarie del quaternione.

Il vettore dell’intensità acustica è dato dalla seguente formula (pressione del suono per velocità della particella), la quale è possibile riscriverla tenendo conto della scomposizione fatta dall’ambisonics, sostituendo i termini in modo corretto. La formula risultante tiene conto anche di diversi fattori tra cui la densità dell’aria e la velocità del suono.

**Slide**

È possibile dividere la trasformata di fourier dell’intensità acustica nelle sue due componenti: parte reale e parte immaginaria.

La parte reale prende il nome di intensità attiva, può essere vista come la media nel tempo dell’intensità acustica e abbiamo già detto che è legata direttamente con la doa (infatti in ambienti senza eco la doa è data proprio dalla direzione opposta a quella dell’intensita attiva),

mentre la parte immaginaria prende il nome di intensità reattiva, indica quando una certa frequenza è dominata da un suono diretto di una singola fonte, o al contrario, quando sono presenti due o più fonti sovrapposte o c’è presenza di eco.

**Slide**

L’input della rete è costitiuto dalle intensità attiva e reattiva, rappresentate in forma quaternionica in questo modo.

Spesso è utile normalizzare queste quantità attraverso un fattore pari alla somma delle densità dell’energia potenziale della pressione del suono e dell’energia cinetica delle particelle.

In questo modo siamo stati in grado di costruire l’input della rete.

Queste quantità sono infatti quelle che vengono passate al primo layer della rete, per essere processate.

Tutto quello che riguarda la struttura della rete stessa ora verrà illustrato da Sveva.

**Slide**