

Raincloud plots Sonification: Uno studio sperimentale

Gaspare Casano, Danilo Dolce

Sommario

Recentemente sono stati introdotti in letteratura i Raincloud plots, un potente strumento statistico per la visualizzazione dei dati. Nel nostro lavoro di studio, effettuiamo la sonificazione di questo plot e sottoponiamo alcuni utenti a delle prove sperimentali il cui risultato è descritto in questo articolo.

1 Introduzione

Negli ultimi anni è nata l'esigenza di disporre di approcci statistici più solidi e trasparenti per la visualizzazione dei dati. Per far fronte a questa necessità, di recente sono stati introdotti i Raincloud plots. Questi "grafici a pioggia" permettono la visualizzazione di dati grezzi, densità di probabilità e statistiche riassuntive chiave come la mediana, la media e intervalli di confidenza pertinenti in un formato più flessibile e accattivante con la minima ridondanza. In questo articolo, preliminarmente, definiamo in maniera più dettagliata questi grafici e successivamente mostriamo il nostro lavoro sperimentale. In particolare, quest'ultimo si divide in due parti distinte:

1. Sonificazione dei Raincloud plots attraverso l'uso di diversi dataset;
2. Esperimenti: sottoporre alcuni utenti all'ascolto delle sonificazioni prodotte al fine di far disegnare loro ciò che hanno ascoltato.

La sonificazione è stata realizzata attraverso i linguaggi di programmazione Pure Data e Python.

Riportiamo tutto il materiale prodotto, quali codice, patch in Pure Data, immagini e file inerenti al progetto nel seguente link: https://github.com/gascas11/Raincloud_Sonification.git

2 Raincloud plots: un nuovo strumento statistico per la visualizzazione dei dati

La visualizzazione efficace dei dati è la chiave per l'interpretazione e la comunicazione dell'analisi dei

dati. Idealmente un grafico statistico dovrebbe bilanciarne la funzionalità, l'interpretabilità e la complessità, il tutto senza sacrificarne l'estetica. Tra i metodi di visualizzazione di punti di dati grezzi vi è il barplot (o grafico a barre). Quest'ultimo utilizza delle barre orizzontali (o linee) per rappresentare la media o la mediana di alcune condizioni e delle barre di errore per rappresentare l'incertezza sul parametro stimato (Figura 1A). Questo approccio negli anni ha subito diverse critiche, tra queste quella di essere soggetto a distorsione e di poter oscurare le differenze nelle distribuzioni.

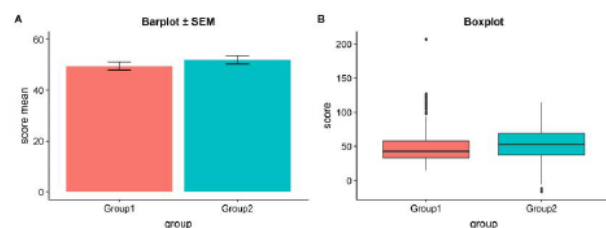


Figura 1: Esempio di a) barplot e b) boxplot tra due gruppi

Per ovviare a queste carenze sono stati proposti vari approcci di visualizzazione illustrati nella Figura 2. In particolare, la Figura 2A rappresenta un semplice miglioramento dato dalla sovrapposizione di singole osservazioni accanto al formato standard del barplot con un grado di jitter randomizzato per migliorarne la visualizzazione. Un approccio complementare statisticamente più robusto a quest'ultimo è il boxplot (Figura 1B) che mostra una mediana lungo l'intervallo interquartile. Nella Figura 2B osserviamo invece dei dot plots (grafici a punti) di due gruppi differenti, che possono essere usati per combinare una visualizzazione simile ad un istogramma con singole osservazioni dei dati. Quando vengono utilizzate statistiche parametriche, spesso è consigliato tracciare la distribuzione delle osservazioni per ricavare importanti informazioni quali, l'aumento della asimmetria o la forma generale di una distribuzione. In tale caso è spesso preferito il "violin plot". In quest'ultimo, dove viene rispecchiata la funzione di densità di probabilità delle osservazioni, combi-

nato con boxplots sovrapposti (Figura 2C) fornisce sia una valutazione sulla distribuzione dei dati e sia l'inferenza statistica a colpo d'occhio (SIG). Tuttavia, dal punto di vista statistico, l'utilizzo del "violin plot" viola la filosofia di minimizzazione del "Data-ink ratio" (ovvero il rapporto dati-inchiostro).

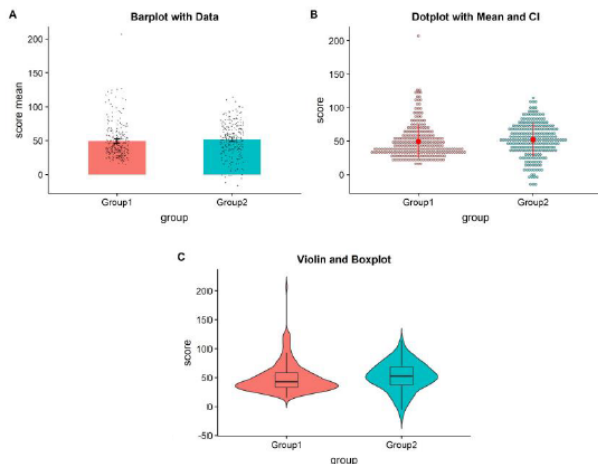


Figura 2: Approcci esistenti per la visualizzazione dei dati

Una soluzione a tutti questi problemi è l'uso dei Raincloud plots illustrati nella Figura 3.

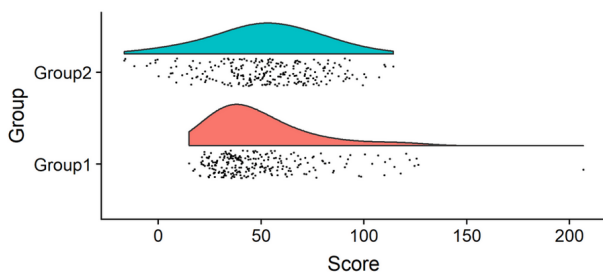


Figura 3: Raincloud plots

I Raincloud plots sono stati introdotti recentemente in [2] come nuovo strumento per la visualizzazione dei dati. Questi "grafici a pioggia" sfruttano diversi modelli di visualizzazione e precursori simili già utilizzati in pubblicazioni esistenti. In particolare, questi combinano un "mezzo violin plot", dei punti di dati grezzi con jitter e una visualizzazione standard della tendenza centrale (ad esempio media o mediana) ed errore, attraverso un boxplot. A seconda dell'analisi in corso, è possibile sostituire la funzione di densità di probabilità (PDF) con opzioni più avanzate come la funzione di densità di probabilità posteriore o altre stime di parametri. Pertanto, i Raincloud plots offrono all'utente la massima utilità e flessibilità garantendo che nulla è "nascosto" e che il lettore abbia tutte le informazioni necessarie per valutare i dati, la

loro distribuzione e l'adeguatezza di eventuali test statistici, il tutto in un formato più accattivante. La Figura 4, mostra come i Raincloud plots possano rivelare informazioni che su un boxplot con dati grezzi potrebbero nascondersi, come una distribuzione bimodale.

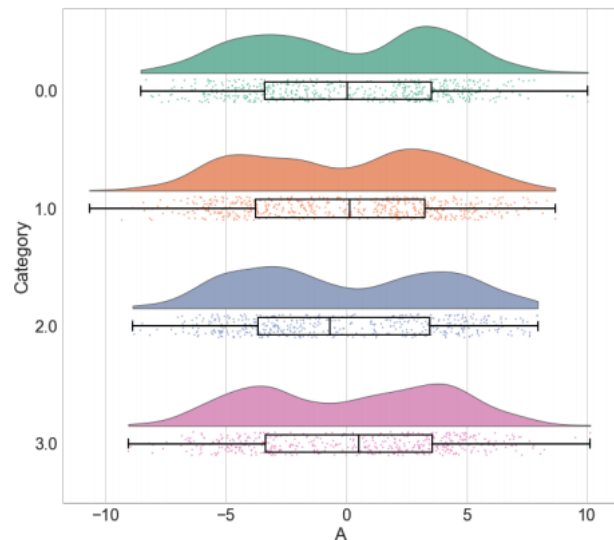


Figura 4: I Raincloud plots non lasciano nulla al caso, forniscono informazioni sia sui singoli punti di dati grezzi che sull'intera distribuzione. Come possiamo notare qui, anche un boxplot con dati grezzi può nascondere aspetti cruciali dei dati come ad esempio la bimodalità.

3 Progettazione del suono e Sonificazione

L'obiettivo primario del nostro lavoro è quello di effettuare delle sonificazioni di diversi Raincloud plots generati a partire da uno specifico set di dati. Queste sonificazioni ci serviranno come base per il nostro studio sperimentale. La sonificazione è stata definita in diversi modi, Hermann l'ha definita come un suono che riflette: proprietà o relazioni oggettive nei dati di input, che ha una sistematica e riproducibile trasformazione in suono e può essere utilizzata con differenti dati di input. Per riprodurre una sonificazione consona alla rappresentazione grafica del Raincloud abbiamo scelto di simulare i suoni del vento e della pioggia rispettivamente per il "cloud" (ovvero la funzione di densità di probabilità) e per la "rain" (ovvero i puntati di dati grezzi). La progettazione di questi suoni è stata realizzata attraverso il linguaggio di programmazione visuale Pure Data. In particolare, per simulare i suoni del vento e della pioggia, abbiamo considerato e modificato rispettivamente le patch riportate nelle Figure 5 e 6 del libro [1].

La tecnica che abbiamo utilizzato per effettuare la sonificazione è la *mappatura dei parametri*, tecnica che consiste nell'adattamento dei dati sugli elementi del suono che deve essere rappresentato.

L'intero processo di sonificazione, da noi messo in

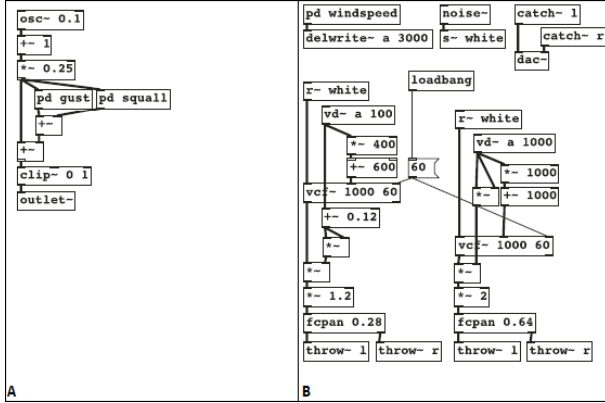


Figura 5: Patch in Pure Data per la simulazione del vento. La Figura 5A mostra la patch windspeed necessaria alla patch della Figura 5B

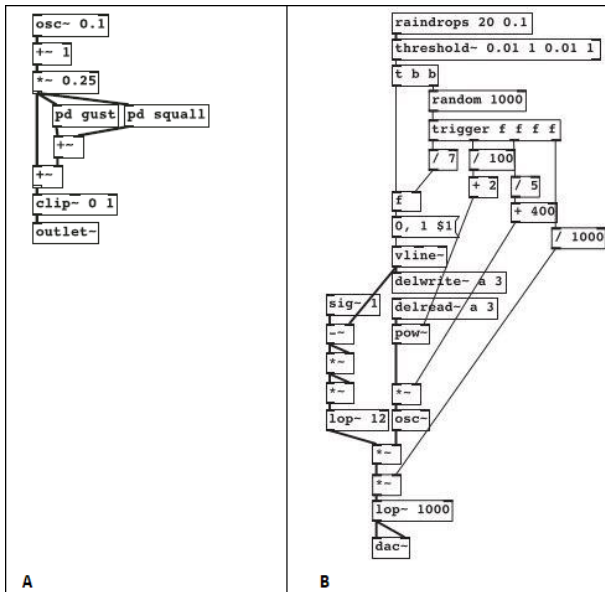


Figura 6: Patch in Pure Data per la simulazione della pioggia. La Figura 5A mostra la patch raindrops necessaria alla patch della Figura 6B

atto, può essere sintetizzato nei seguenti 4 punti:

1. Lettura di uno specifico dataset;
2. Generazione del Raincloud plot associato al set di dati caricato;
3. Invio dei dati necessari per la sonificazione a Pure Data;
4. Ricezione dei dati e riproduzione in tempo reale della sonificazione con PureData;

I precedenti punti sono stati implementati con il supporto del linguaggio di programmazione Python e gli IDE Spyder e Jupyter. Per l'invio dei dati in tempo reale è stato necessario creare un'apposita socket che svolge il suo lavoro in locale mediante il protocollo UDP. Mentre, per la ricezione dei dati in byte e la loro conversione nel tipo di dato appropriato in Pure Data, ci siamo avvalsi di alcuni pacchetti esterni, quali rispettivamente *mrpeach* e *moocow*. La Figura 7 mostra gli oggetti *udpreceive* e *bytes2any* utilizzati in Pure Data rispettivamente per ricevere e convertire i dati nel tipo di formato appropriato. Quello che vogliamo evidenziare inoltre è che l'invio dei dati a Pure Data per la relativa sonificazione avviene mediante una lettura del Raincloud plot in esame, che procede da sinistra verso destra. Le varie sonificazioni realizzate possono essere consultate al seguente link: https://github.com/gascas11/Raincloud_Sonification/tree/master/Raincloud_Project_Source

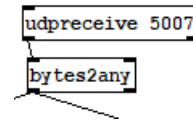


Figura 7: Oggetti utilizzati per la ricezione dei dati necessari per la sonificazione in Pure Data. L'oggetto *udpreceive* permette di ricevere i dati tramite la porta specificata attraverso l'uso del protocollo UDP, mentre l'oggetto *bytes2any* effettua la loro conversione nel tipo di dato appropriato.

4 Sonificazione di un Raincloud plot con boxplot

Il nostro lavoro oltre ad includere una parte dedicata agli esperimenti, include una parte dedicata alla progettazione e alla realizzazione di varie sonificazioni di Raincloud plots, queste generate attraverso l'uso di diversi dataset. In particolare, oltre a realizzare la sonificazione di semplici Raincloud plots, abbiamo progettato e realizzato una sonificazione più particolare che prevede la combinazione tra il Raincloud plot e il boxplot relativo allo stesso set di dati. La combinazione di un Raincloud plot con un boxplot offre un'ulteriore panoramica dei dati grezzi e della distribuzione delle probabilità tramite l'utilizzo della mediana e degli intervalli di confidenza (o quartili). Per realizzare questa sonificazione abbiamo utilizzato la patch della sonificazione presa in considerazione

negli esperimenti (vedi Par. 5.1) e creato due ulteriori patch in Pure Data per la riproduzione di un beep breve per la rappresentazione sonora dei quartili e un beep leggermente più lungo per la mediana. La Figura 8 mostra il Raincloud plot con boxplot di cui abbiamo effettuato la sonificazione, mentre le Figure 9 e 10 mostrano rispettivamente le patch in Pure Data costruite per la riproduzione del beep dei quartili e della mediana.

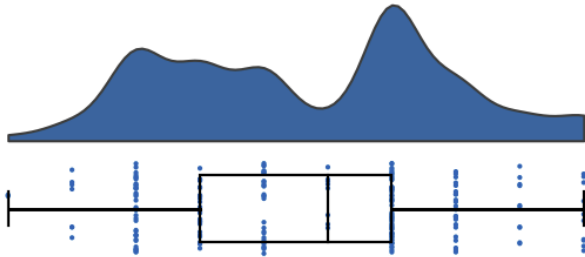


Figura 8: Raincloud plot con boxplot sonificato

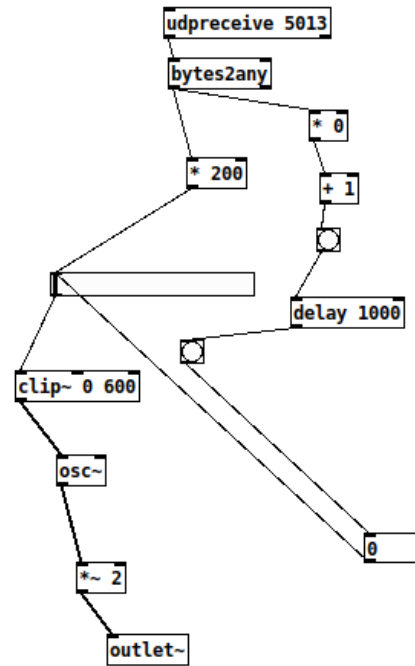


Figura 10: Patch Pure Data per la sonificazione della mediana

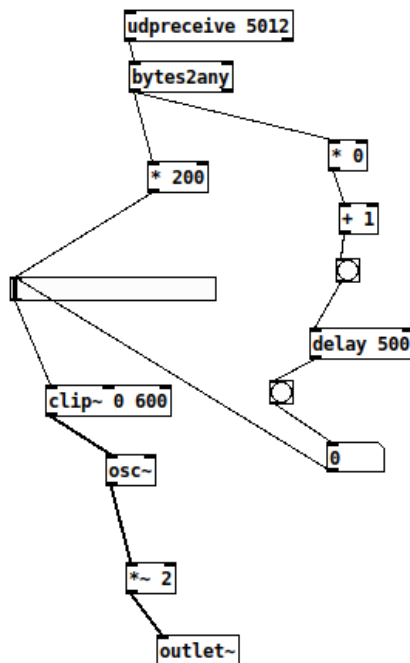


Figura 9: Patch Pure Data per la sonificazione dei quartili

5 Esperimenti e discussione dei risultati

Un aspetto affascinante legato allo studio dell'esplorazione sensoriale dei dati è poter sperimentare nuove tecniche o soluzioni che permettono di ottenere delle migliorie o dei benefici per chi successivamente li dovrà realmente interpretare. Il nostro studio sperimentale si concentra nello specifico sulla percezione uditiva della sonificazione di

un Raincloud plot multimodale (cioè con più gobbe). In particolare, il nostro esperimento consiste nel sottoporre ad alcuni soggetti, undici in totale e con un range di età che va dai 20 ai 60 anni, una prova che prevede la rappresentazione grafica di questa sonificazione, che ricordiamo essere l'unione di due suoni: il vento e la pioggia, elementi caratteristici di un Raincloud plot. Prima di eseguire ciascun esperimento, ogni soggetto è stato edotto circa ciò che rappresenta un Raincloud plot.

5.1 Strumenti per l'esecuzione degli esperimenti e parametri di valutazione adottati

Per sottoporre i soggetti a questa prova sperimentale abbiamo realizzato un modello specifico che consiste in un foglio opportunamente millimetrato in formato A4. Questo foglio, disponibile al seguente link: https://github.com/gascas11/Raincloud_Sonification/blob/master/Raincloud_Project_Source/Esperimenti/Modello_Esperimento.pdf, prevede l'inserimento di alcuni dati relativi al soggetto, quali nome, cognome e data di esecuzione dell'esperimento. Inoltre, al centro di quest'ultimo abbiamo disegnato i bordi di un apposito rettangolo che delimitano l'area normalizzata della parte superiore del Raincloud plot di riferimento (ovvero le "nuvole" o "clouds") in cui i soggetti

andranno a disegnare una curva corrispondente a ciò che percepiscono dalla sonificazione.

La sonificazione che abbiamo adottato per i nostri esperimenti è riportata al seguente link: https://github.com/gascas11/Raincloud_Sonification/blob/master/Raincloud_Project_Source/Sonificazione_Multimodale/Sonificazione_Multimodale.wav. La Figura 11 mostra la parte superiore del Raincloud plot di riferimento scelta per l'esecuzione degli esperimenti.

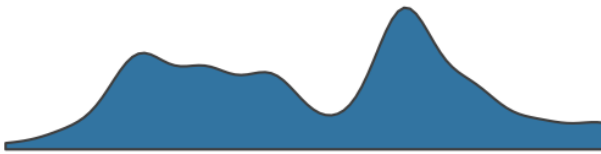


Figura 11: Clouds del Raincloud di riferimento

Per effettuare la valutazione delle singole prove, che consiste nel confrontare la curva disegnata dal soggetto con la curva corrispondente al Raincloud plot di riferimento, abbiamo considerato le singole distanze tra tutti i punti delle due curve per poi andare a ricavare la somma dei minimi quadrati di quest'ultime.

5.2 Discussione dei risultati

Quello che ci aspettavamo dai risultati finali non è certamente che tutti i soggetti siano in grado di rappresentare al meglio la curva del Raincloud plot di riferimento, visto e considerato che la rappresentazione grafica a mano libera di una curva a partire da un suono non è del tutto semplice e precisa, ma che almeno il 30% di quest'ultimi riuscisse a disegnare circa ciò che realmente la sonificazione vuole comunicare. In termini pratici, una curva che cresce fino ad un certo punto, decresce per poi risalire fino al picco ed infine ridursi fino ad azzerarsi. La Tabella 1 mostra i risultati finali per ciascun soggetto, corrispondenti alle somme dei minimi quadrati tra le distanze della loro rappresentazione e quella del Raincloud di riferimento. I risultati sono ordinati dal punteggio più basso al più alto, dove con punteggio basso si intende il valore più piccolo prossimo allo zero che indica che la curva rappresentata dal soggetto è meno distante dal Raincloud plot di riferimento, viceversa con punteggio alto. La Figura 12 mostra l'esperimento che ha ottenuto il miglior punteggio in termini di distanza rispetto alla curva del Raincloud plot di riferimento. Come

Nome soggetto	Risultato
Part. 1	153.8
Part. 2	275.55
Part. 3	293.49
Part. 4	294.85
Part. 5	304.89
Part. 6	458.33
Part. 7	489.15
Part. 8	490.39
Part. 9	618.12
Part. 10	621.12
Part. 11	1249.61

Tabella 1: Risultati degli undici esperimenti eseguiti. I valori riportati per ciascun soggetto rappresentano le somme dei minimi quadrati delle distanze tra la rappresentazione del soggetto e quella originale.

possiamo osservare, il Partecipante 1, ascoltando la sonificazione è riuscito a rappresentare una curva (linea nera) che si avvicina molto a quella originale (linea blu) anche se non sembra ricalcare al meglio la decrescita della prima gobba e la crescita che procede fino al picco. A differenza

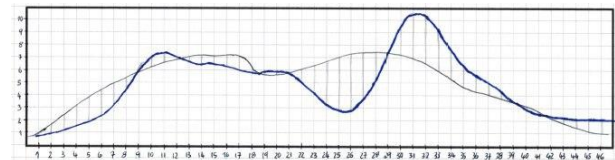


Figura 12: Esperimento Part. 1

dell'esperimento della Figura 12 che ha ottenuto il miglior punteggio in termini di distanza, ma che non riesce ad evidenziare con esattezza l'andamento della curva originale, riteniamo che l'esperimento del Partecipante 2 riportato nella Figura 13, che ha ottenuto il secondo punteggio migliore, sia il più soddisfacente dal punto di vista rappresentativo. Infatti, la curva disegnata dal soggetto (linea nera) rispecchia quello che la sonificazione vuole comunicare. Nello specifico possiamo notare l'andamento crescente iniziale della curva, la sua decrescita per poi ricrescere fino al picco e la sua riduzione fino al suo azzeramento. Per i restanti nove esperimenti non mostreremo la figura corrispondente, queste sono reperibili al seguente link: https://github.com/gascas11/Raincloud_Sonification/blob/master/Raincloud_Project_Source/Esperimenti/Esperimenti.pdf. Nonostante ciò espletiamo per ognuno di essi un commento:

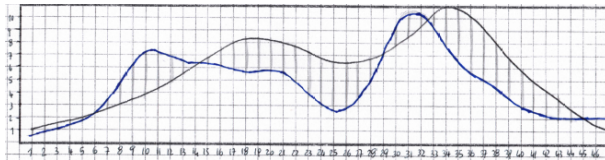


Figura 13: Esperimento Part. 2

- Part. 3: la curva disegnata è molto fedele alla reale sonificazione. Il soggetto riesce a percepire le due gobbe del Raincloud plot di riferimento ma le pone allo stesso livello di ampiezza non evidenziando l'ampiezza maggiore che si avrebbe in realtà con la seconda gobba rispetto alla prima.
- Part. 4: la curva disegnata riesce a rappresentare circa l'ampiezza massima che viene raggiunta con il Raincloud plot di riferimento ma non evidenzia la prima gobba e la sua decrescita rispetto all'originale. **Nota:** il soggetto, dopo aver effettuato l'esperimento, alla richiesta di riascolto della sonificazione e individuazione di possibili errori nella rappresentazione, riesce ad identificare perfettamente ciò che la sonificazione vuole comunicare.
- Part. 5: la curva disegnata è quasi fedele alla reale sonificazione. Il soggetto riesce a percepire le due gobbe del Raincloud plot di riferimento, ma le pone allo stesso livello di ampiezza non evidenziando l'ampiezza maggiore che si avrebbe in realtà con la seconda gobba rispetto alla prima.
- Part. 6: la curva disegnata rispecchia in parte la reale sonificazione. In particolare il soggetto riesce a percepire le due gobbe ma le inverte nella loro rappresentazione grafica.
- Part. 7: la curva disegnata ha un andamento iniziale e finale che rispecchia la curva del Raincloud plot di riferimento. Il soggetto non riesce però a rappresentare la decrescita della prima gobba e vede quest'ultima con un andamento costante.
- Part. 8: la curva disegnata ha un ottimo andamento iniziale, nonostante ciò non riesce ad identificare la prima gobba e di conseguenza la sua decrescita. Il decremento di quella che dovrebbe rappresentare la gobba con ampiezza massima risulta essere leggermente più lungo.
- Part. 9: la curva disegnata ha un andamento iniziale discreto, tuttavia essa risulta esse-

re gradualmente crescente quasi per tutta la sua durata con un piccolo decremento e un successivo incremento che vede poi una sua riduzione finale.

- Part. 10: la curva disegnata ha un discreto andamento iniziale e un ottimo andamento finale. Il soggetto però non riesce a rappresentare correttamente le due gobbe vedendo il tutto come una curva che cresce fino al picco e decresce.
- Part. 11: la curva disegnata si discosta molto dalla curva del Raincloud plot di riferimento.

6 Conclusione

I Raincloud plots come abbiamo visto sono un ottimo strumento per la visualizzazione dei dati. Il nostro lavoro sperimentale aveva l'obiettivo di studiare e verificare ciò che i soggetti riuscissero a percepire dalla sonificazione di un Raincloud plot, questo per capire se quest'ultima possa fornire un ulteriore beneficio nell'interpretazione dei dati. Da quello che si evince dagli esperimenti effettuati, circa tre soggetti su undici riescono a fornire una rappresentazione che si avvicina alla reale sonificazione proposta. Tuttavia anche se i restanti otto soggetti non riescono a riprodurre al meglio la sonificazione, ulteriori feedback rilasciati mediante questionari dimostrano come quasi la metà di essi riesce a percepire l'andamento di quest'ultima e che spesso risulta difficile o poco preciso rappresentare a mano libera ciò che si ascolta. In conclusione, quello che possiamo affermare è che la sonificazione risulta essere uno strumento utile per l'ausilio all'interpretazione dei dati. Un ulteriore direzione futura di sviluppo di questo esperimento potrebbe essere quella di potenziare e generalizzare il software per la sonificazione dei Raincloud plots.

7 Bibliografia

- [1] Farnell A., Designing Sound, London, The MIT Press, 2010
- [2] Allen M. - Poggiali D. - Whitaker K. - Marshall T. - Kievit R., Raincloud plots: a multi-platform tool for robust data visualization, in "Wellcome Open Research", a.MMXIX, n.1, Aprile 2019