Analyzing brainstem auditory responses in Alzheimer using order restricted methods

Xinzhu Liu

Abstract

Two kinds of order restricted methods, isotonic regression and taut strings are used in this paper to model brainstem auditory responses in Alzheimer disease. It is shown that the two methods present similar result but taut string method is faster in practical use. Residual bootstrap technique is also used to help analyze features. Order restricted methods are versatile and help in the identification of main features in the responses that can help in discriminating between Alzheimer's patients and healthy individual.

1 Introduction

1.1 Alzheimer disease

Alzheimer disease (AD), is the most common form of dementia. The disease is named after a German psychiatrist and neuropathologist Alois Alzheimer, who first reported a death case of this disease in 1906. Before that, the ancient Roman and Greek philosophers have described old patients increasing dementia. Most often, Alzheimer is diagnosed in people over 65 years of age, although the less-prevalent early-onset Alzheimer can occur much earlier [5].

Early symptoms of Alzheimer are often mistakenly considered to be 'age-related' concerns, or stress [8]. Most often, patients have difficulty in remembering recent events after onset of the disease. As the disease advances, symptoms can include confusion, trouble with language, mood swings, and long-term memory loss. In the end, patients lose vital bodily functions, ultimately leading to death. Alzheimer disease develops for an unknown and variable amount of time before becoming fully apparent, and it can progress undiagnosed for years. On average, the life expectancy following diagnosis is approximately seven years [13]. After diagnosis fewer than three percent of

individuals live more than fourteen years [14].
The cause and progression of Alzheimer's disease are still unknown. Research shows that the disease is associated with plaques and tangles in the brain [15]. When AD is suspected, tests that evaluate behavior and mental abilities are usually used to confirm the diagnosis, often followed by a brain scan if available.

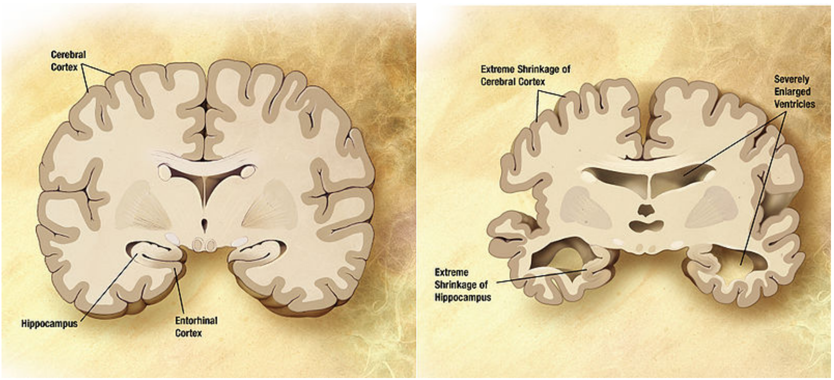

Figure 1: Comparison of a normal aged brain (left) and the brain of a person with Alzheimer's (right). Differential characteristics are pointed out. (Taken from http://en.wikipedia.org/wiki/Alzheimer)

There is still no known cure for the Alzheimer disease. Current treatments only help with the symptoms of the disease. There are also no available treatments that stop or reverse the progression of the disease. As of 2012, more than 1000 clinical trials have been or are being conducted to find ways to treat the disease, but it is unknown if any of the tested treatments will work. There is no conclusive evidence supporting an effect, although exercise, mental stimulation and a balanced diet have been suggested as methods to delay symptoms in healthy older individuals.
In 2006, there were 26.6 million patients suffering worldwide. And it is predicted that Alzheimer is going to affect 1 in 85 people globally by 2050 [4]. Alzheimer disease may be among the most costly diseases not only for society in Europe and the United States, but also in south American countries such as Argentina, or Asian countries such as South Korea. These costs will probably increase with the ageing of society, becoming an important social problem. Numbers vary between studies but dementia costs worldwide have been calculated around 160 billion US dollars [18], while costs of Alzheimer in the United States may be 100 billion dollars each year [12].
Since Alzheimer disease is an important problem around the whole world, looking for difference between patients and healthy people in auditory brain-

stem responses may leads to a new direction to find the cause and cure of Alzheimer disease.

1.2 Auditory brainstem response

The auditory brainstem response (ABR) is an objective technique that measures the electrical activity of the subcortical nerve cells along the auditory pathway upon auditory stimulation [10]. The auditory brainstem response usually contains seven positive peaks and is recorded within 10 ms following stimulation onset. This technique's advantages include that it does not depend on the level of consciousness nor on the cooperation of the subject and it is not affected by general anesthetics [6].

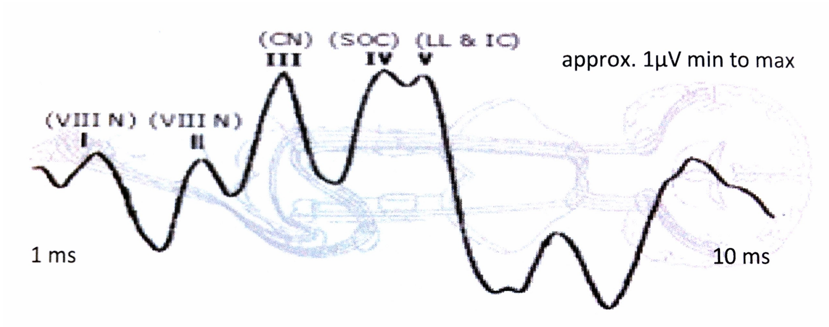

Figure 2: The auditory brainstem response waves

2 Data and Problem

The data set we use in this paper is provided by the Sensodetect company. We compare data of 9 Alzheimer's patients and 7 reference persons. Each individual has 2 files. One is a recording from left ear and the other is recorded from right ear. Each file contains a 1283×256 real-valued matrix which record ABR to 1283 clicks within 10ms.

Type	Female	Male	Total
ALZ	3	6	9
REF	3	4	7
Total	6	10	16

Table 1: Frequency counts for patient group and reference group (each individual has 2 ears' data)

The patient group and reference group contain individuals from different age and gender.

Gender	Type	Year of birth
Female	ALZ	1925, 1941, 1942
Female	REF	1935, 1940, 1944
Male	ALZ	1931, 1936jor, 1936, 1937, 1939, 1946
Male	REF	1934, 1937, 1940, 1944

Table 2: details about data

There are two male patients born in 1936. The younger one is labeled as "36jor". So the figure named "F25 Left ALZ" describe the result of the left ear data of a female patient born in 1925. And a picture named "M37 Right REF" shows the result of the right ear data of a male born in 1937 in reference group.
A main problem we discuss in this thesis is modeling ABR data of Alzheimer's patient and reference group using order restricted methods. We want to use nonparametric statistical methods to classify and characterize the patient group against the healthy controls based on the column mean of the data. An example of the column mean is shown in figure 3. The ultimate goals are to find the cause of the disease and contribute to an effective cure and to prevention.

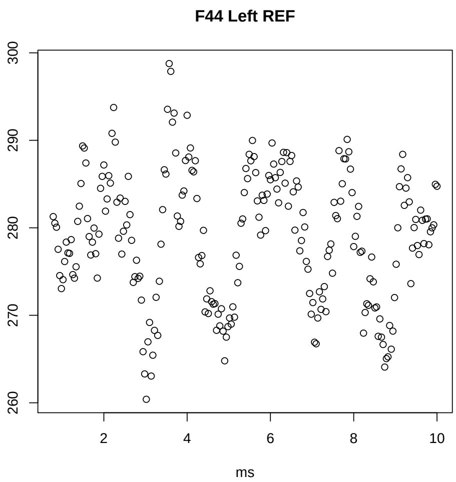

Figure 3: The column mean of the data

3 Theory and Method

A nonparametric problem means that we want to estimate a function f which lies in a function set $\mathcal{F}$ which is infinite-dimensional. It is the opposite of parametric problem. $\mathcal{F}$ is not necessary to be closed, and individual variables can be assumed to belong to parametric distributions.

Example 1. *An nonparametric inference problem: given regression data (t_i, y_i), for $i = 1, \ldots, n$, with $y_i = f(t_i) + \epsilon_i$. The problem is to estimate f when f lies in $\mathcal{F}$. $\mathcal{F}$ could be:*

1. $\mathcal{F} = \{f : \mathbb{R} \to \mathbb{R} \quad \textit{such} \quad \textit{that} \quad f \in \mathbb{C}^2\}$.
2. $\mathcal{F} = \{f : \mathbb{R} \to \mathbb{R} \quad \textit{such} \quad \textit{that} \quad f \quad \textit{decreasing}\}$.
3. *For density estimation,* $\mathcal{F} = \{m : [0,1] \to \mathbb{R} \quad m \quad \textit{decreasing}\}$

We have data set(t_i, y_i), our problem is to find an estimate $\hat{f}$ of the unknown function $f(t_i)$ in the model

$$y_i = y(t_i) = f(t_i) + \varepsilon_i \tag{1}$$

where ε_i are identically distributed random variables with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$, and the unknown $f \in \mathcal{F}$. $\mathcal{F}$ is modeled as the set of all nonparametric multimode functions with unknown number of mode, and unknown position of peaks and troughs.

3.1 Kernel smoothing

A kernel smoother is a statistical technique for estimating a real valued function f(t), by using data (t_i, y_i) from (1), when no parametric model for this function is known. The estimated function is smooth, and the level of smoothness is set by a single parameter.
Let $K_d(t_0, t)$ be a kernel defined by

$$K_d(t_0, t) = D\left(\frac{\|t - t_0\|}{d(t_0)}\right)$$

where:
$t, t_0 \in \mathbb{R}$, $\|\cdot\|$ is the Euclidean norm, $d(t_0)$ is a bandwidth, $D(\cdot)$ is a positive real valued function, called kernel function, which value is decreasing (or not increasing) for the increasing distance between the t and t_0.
Let $\hat{y}(t) : \mathbb{R} \to \mathbb{R}$ be a continuous function of t. For each $t_0 \in \mathbb{R}$, the

Nadaraya-Watson kernel-weighted average (smooth $y(t)$ estimation) is defined by

$$\hat{y}(t_0) = \frac{\sum_{i=1}^{N} K_d(t_0, t_i) y(t_i)}{\sum_{i=1}^{N} K_d(t_0, t_i)}$$

where N is the number of observed points, $y(t_i)$ are the observations at t_i points.

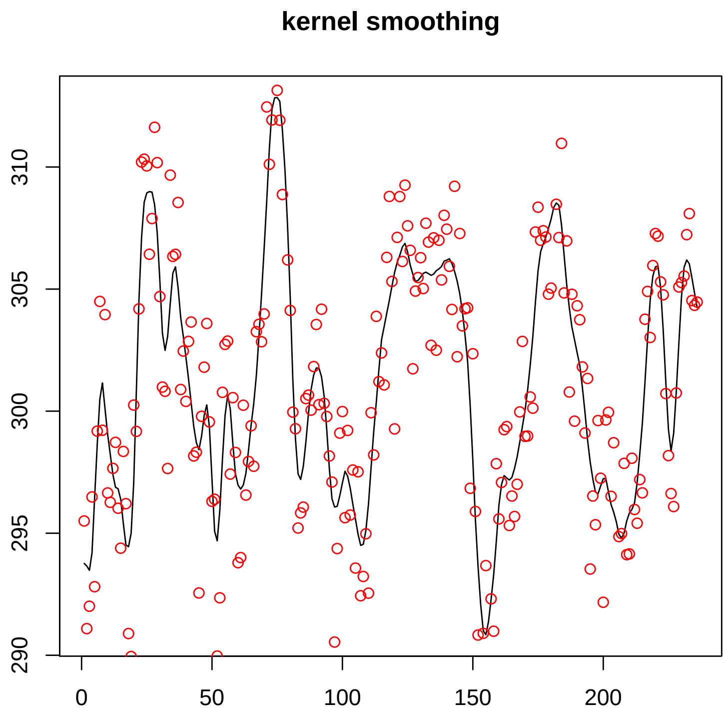

Figure 4: kernel smoother using Gaussian kernel

Epanechnikov kernel, Tri-cube kernel and Gaussian kernel are popular used for smoothing.

kernel type	K(u)
Epanechnikov	$\frac{3}{4}(1-u^2)\ 1_{(\lvert u\rvert\leq 1)}$
Tri-cube	$\frac{70}{81}(1-\lvert u\rvert^3)^3\ \mathbf{1}_{\{\lvert u\rvert\leq 1\}}$
Gaussian	$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$

Table 3: popular kernel functions

3.2 Isotonic regression

One way to deal with multimode problem which the number of mode and the positions of peaks and troughs are known is separating it to several unimodes, then treat the unimode as two isotonic functions. This section derives characterizations for the solution to the isotonic regression and presents algorithms for the calculation of the solution. The main reference for all present results in following two subsection is [16].

3.2.1 Theory

In order to define an isotonic function, we need to talk about order relations first. We study functions defined on a set T taking real values, on which is defined an order relation $\leq$.

Definition 1. *A binary relation $\leq$ on T is called a quasi-order if for all s, t, u in T*

$$\textit{(i) } t \leq t,$$

$$\textit{(ii) } s \leq t, t \leq u, \Rightarrow s \leq u,$$

A quasi-order is called a partial order if it further

$$\textit{(iii) } s \leq t, t \leq s, \Rightarrow s = t,$$

A partial order is called a simple order if also

(iv) any two elements s, t in T are comparable, i.e. either $s \leq t$, or $t \leq s$.

The following examples explain the definitions of orders clearly.

Example 2. *1. The ordinary order $\leq$ on $\mathbb{R}$ is a simple order.*

2. The order on $\mathbb{R}^2$ defined by $s \leq t$ for $s = (s_1, s_2)$, $t = (t_1, t_2)$ if $s_1 \leq t_1$ and $s_2 \leq t_2$ is a partial order. Because we know order relations in two axises direction but we cannot compare s, t in any other direction in figure 5.

3. The order $<$(strictly less than) on $\mathbb{R}$ is a quasi-order.

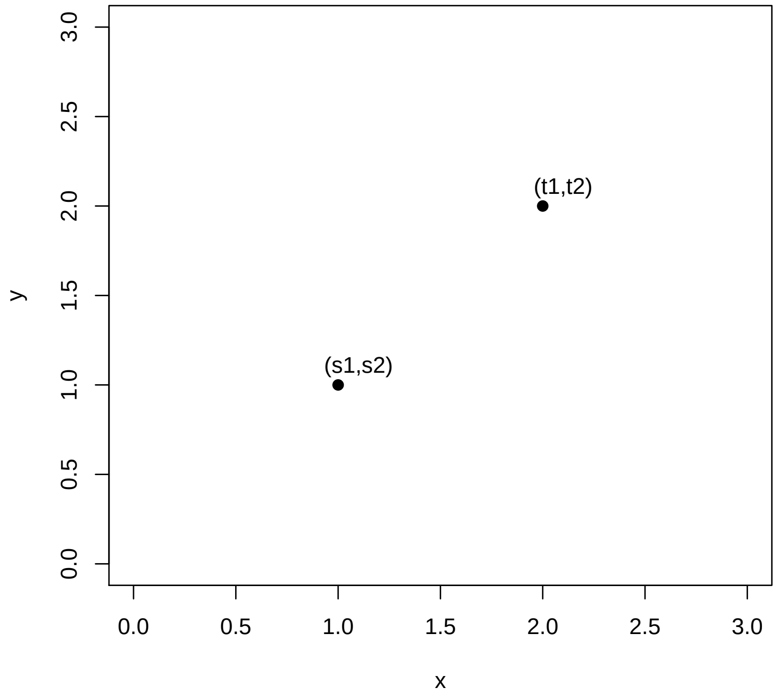

Figure 5: example 2.2: partial order

Now we can define an isotonic function given a quasi-order.

Definition 2. *If we have given a quasi-order on T we say that a realvalued function f defined on T is isotonic, if $s \leq t$ implies that $f(s) \leq f(t)$. The class of all isotonic functions on T is denoted by $\mathcal{F}$.*

We can look at isotonic as a substitute for monotonically increasing.
Assume T is a finite set with a quasi-order $\leq$ defined on it, and assume we have obtained measurements (t_i, y_i) for $t_i \in T$ such that

$$y_i = y(t_i) = f(t_i) + \epsilon_i$$

where ϵ_i can be seen as the measurements errors. We want to define a way to obtain an estimate of f if we know that f is an isotonic function satisfied common conditions stated above.

Definition 3. *If y is any function on T, and $\{w_i\}_{t_i \in T}$ are given weights, then we define the isotonic regression $\hat{f}$ of y as*

$$\hat{f} = \mathrm{argmin}_{z \in \mathcal{F}} \sum_{t_i \in T} (y(t_i) - z(t_i))^2 w_i$$

□

In our problem, it is not clear how to care weights, so the solution of the isotonic regression problem is

$$\hat{f} = \mathrm{argmin}_{z \in \mathcal{F}} \sum_{t_i \in T} (y(t_i) - z(t_i))^2$$

Before deriving an algorithm for $\hat{f}$, we need to show some properties of isotonic functions first.
Our optimization problem is that we want to minimize a convex function $\varphi(z) = \sum_{t_i \in T} (y_i - z(t_i))^2$ over the set $\mathcal{F}$ of isotonic functions. The conditions of the set $\mathcal{F}$ are listed in the following lemma.

Lemma 1. *Assume that T is a finite set. Then the set $\mathcal{F}$ is closed, convex cone, i.e.*

1. $z_1, z_2 \in \mathcal{F}, p \in (0,1) \Rightarrow pz_1 + (1-p)z_2 \in \mathcal{F}$, *(convexity)*
2. $z \in \mathcal{F}, p > 0 \Rightarrow pz \in \mathcal{F}$, *(cone)*
3. $\{z_i\}_{i \geq 1} \in \mathcal{F}, \|z_n - z\| \to 0$, *as* $n \to \infty \Rightarrow z \in \mathcal{F}$*(closedness).*

Thus we want to minimize a convex function over a closed, convex set.

Theorem 1. *If $\mathcal{F}$ is any convex set of functions on the set T and y is an arbitrary function on T, then*

$$\hat{f} = argmin_{z \in \mathcal{F}} \sum_{t_i \in T} (y(t_i) - z(t_i))^2 \tag{2}$$

if and only if

$$(y - \hat{f}, \hat{f} - z) = [\sum_{t_i \in T}(y(t_i) - \hat{f}(t_i))(\hat{f}(t_i) - z(t_i))] \geq 0 \qquad (3)$$

for all $z \in \mathcal{F}$. The function $\hat{f}$ is unique if it exists. □

There exists a solution $\hat{f}$ in $\mathcal{F}$ since $\mathcal{F}$ is closed and furthermore since $\sum_{t_i \in T}(y(t_i) - z(t_i))^2$ is strictly convex, the solution should be unique. We illustrate the statement of Theorem 1 in figure 6.

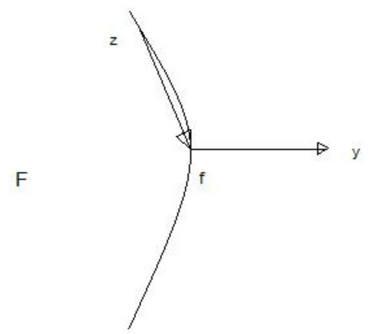

Figure 6: theorem 1

The following theorem shows that $\hat{f}$ always exists.

Theorem 2. *If y, w are given functions of the finite set T the isotonic regression in equation(2) exists.*

To prove Theorem 2, from lemma 1, $\mathcal{F}$ is a closed convex cone, and $\varphi(z) = \sum_{t_i \in T}(y(t_i) - z(t_i))^2$ is a convex function. Therefore $\mathcal{F}$ has a unique minimum, which we denote by $\hat{f}$, in $\mathcal{F}$, and thus the isotonic function $\hat{f}$ exists.

3.2.2 Algorithm

There are two important corollaries of Theorem 1, and they are needed to construct algorithms of isotonic regression problem.

Corollary 1. *If $\mathcal{F}$ is a convex cone of functions on T and y is an arbitrary function on T, then $\hat{f}$ is the solution to inequality 2 if and only if*

$$(y - \hat{f}, \hat{f}) = 0, \tag{4}$$

$$(y - \hat{f}, z) \leq 0, \tag{5}$$

for all $z \in \mathcal{F}$.

Proof. ($\Uparrow$):

$$\begin{aligned} (y - \hat{f}, \hat{f} - z) &= \sum_{t_i \in T} (y(t_i) - \hat{f}(t_i))(\hat{f}(t_i) - z(t_i)) \\ &= \sum_{t_i \in T} (y(t_i) - \hat{f}(t_i))\hat{f}(t_i) - \sum_{t_i \in T} (y(t_i) - \hat{f}(t_i))z(t_i) \\ &= (y - \hat{f}, \hat{f}) - (y - \hat{f}, z) \end{aligned}$$

If the inequalities 4, 5 are satisfied, $(y - \hat{f}, \hat{f} - z) \geq 0$. Then by Theorem 1, $\hat{f}$ is the solution.
($\Downarrow$): Let $z = c\hat{f}$, since inequality 3 holds, then

$$(y - \hat{f}, \hat{f} - c\hat{f}) = (1 - c)(y - \hat{f}, \hat{f}) \geq 0$$

if we apply $c > 1$ and $0 < c < 1$, we will get $(y - \hat{f}, \hat{f}) = 0$.

$$(y - \hat{f}, z) = (y - \hat{f}, \hat{f}) - (y - \hat{f}, \hat{f} - z)$$

Since inequalities 4 and 3 hold, $(y - \hat{f}, z) \leq 0$ for all $z \in \mathcal{F}$. □

Corollary 2. *If $\mathcal{F}$ is a convex cone of functions on T and y is an arbitrary function on T, and assume it is generated by the function $\{\upsilon\}_{i=1}^{n}$.then $\hat{f}$ is the solution to inequality 3 if and only if*

$$(y - \hat{f}, \upsilon_i) \leq 0, \quad for \quad i = 1, \ldots, n, \tag{6}$$

$$(y - \hat{f}, \upsilon_i) = 0, \quad if \hat{\lambda}(i) > 0, \tag{7}$$

where $\hat{\lambda}(i) = \hat{f}(t_i) - \hat{f}(t_{i-1})$.

We can easily prove this corollary by Corollary 1, using

$$(y - \hat{f}, z) = \sum_{i=1}^{n} \lambda(i)(y - \hat{f}, \upsilon(i))$$

when $z = \hat{f}$, $\lambda(i) = \hat{\lambda}(i) = \hat{f}(t_i) - \hat{f}(t_{i-1})$. $\square$

By stating a relation between our observations and the solution to the isotonic regression problem below, we will draw three algorithms. First, we introduce two notations for the cumulative sums:

$$\hat{F}_k = \sum_i^k \hat{f}(t_i)$$

$$Y_k = \sum_i^k y_i$$

Plot the points $\hat{m}_k = (k, \hat{M}_k), k = 1, \ldots, n$, and points $m_k = (k, Y_k), k = 1, \ldots, n$. We can summary our findings in next lemma.

Lemma 2. *The cumulative sums of the observations and the solution to an isotonic regression problem have following relations:*

1. *The function described by the points $\hat{m}_k$ is a continuous convex function.*
2. *The function $\hat{m}_k$ is a minorant to m_k.*
3. *$\hat{m}_k$ and m_k touch at a point where $\hat{m}_k$ is strictly convex.*

Before the proof of this lemma, we need rewrite every $z \in \mathcal{F}$. Assume that there are n points in T, then every $z \in \mathcal{F}$ can be written as

$$\begin{aligned} z(t_k) &= z(t_k) - z(t_{k-1}) + z(t_{k-1}) - z(t_{k-2}) + \ldots + z(t_2) - z(t_1) + z(t_1) \\ &= \sum_{i=2}^{n} (z(t_i) - z(t_{i-1})) + max(z(t_1), 0) + (-min(z(t_1), 0)) \\ &= \sum_{i=1}^{n} \upsilon_i(t_k)\lambda(i) + (-\upsilon_1(t_k))\lambda(-1) \end{aligned}$$

where

$$\upsilon_i(t) = 1_{\{t_i \leq t\}}, for \quad t \in T, i = 1, \ldots, n.$$
$$\lambda(i) = z(t_i) - z(t_i - 1), \quad i = 2, \ldots, n.$$
$$\lambda(1) = max(z(t_1), 0),$$
$$\lambda(-1) = -min(z(t_1), 0).$$

Such that every $z \in \mathcal{F}$ can be generated based on the increasing functions $-\upsilon_1, \upsilon_1, \ldots, \upsilon_n$, and the weights $\lambda(-1), \lambda(1), \ldots, \lambda(n)$ are nonnegative.
Proof.

1. Since $\hat{f}$ is the solution to an isotonic regression problem, $\hat{F}_k - \hat{F}_{k-1} = \hat{f}(t_k)$ is increasing. $\hat{m}_k = (k, \hat{F}_k)$ is continuous and convex.

2. To prove $\hat{F}_k \leq Y_k$ for all $k = 1, \ldots, n$, which means to prove $\sum_i^k (y_i - \hat{f}(t_i)) \geq 0$ for all $k = 1, \ldots, n$.
   Apply Corollary 2 to arbitrary $z(t_k)$.

$$\begin{aligned}(y - \hat{f}, \upsilon_k) &= \sum_{i=1}^{n} (y_i - \hat{f}(t_i))\upsilon_k \\ &= \sum_{i=1}^{n} (y_i - \hat{f}(t_i))1_{\{t_k \leq t_i\}} \\ &= \sum_{i=k}^{n} (y_i - \hat{f}(t_i))\end{aligned}$$

   Apply Corollary 2 to this situation,

$$(y - \hat{f}, \upsilon_k) \begin{cases} \leq 0 & for \quad k = 1, \ldots, n \\ = 0 & if \quad \hat{\lambda}(k) = \hat{f}(t_k) - \hat{f}(t_{k-1}) > 0 \end{cases}$$

   if $k = 1$, since $(y - \hat{f}, \upsilon_1) \leq 0$ and $(y - \hat{f}, -\upsilon_1) \leq 0$, then $(y - \hat{f}, \upsilon_1) = 0$. Since

$$\sum_{i=k}^{n} (y - \hat{f}(t_i)) = (y - \hat{f}, \upsilon_k)$$

   $\sum_i^k (y_i - \hat{f}(t_i))$ can be written as

$$\begin{aligned}\sum_{i}^{k} (y_i - \hat{f}(t_i)) &= (y - \hat{f}, \upsilon_1) - (y - \hat{f}, \upsilon_{k+1}) \\ &= -(y - \hat{f}, \upsilon_{k+1})\end{aligned}$$

   According to our finding,

$$\sum_{i}^{k} (y_i - \hat{f}(t_i)) \begin{cases} \geq 0 & for \quad k = 1, \ldots, n \\ = 0 & if \quad \hat{\lambda}(k) = \hat{f}(t_k) - \hat{f}(t_{k-1}) > 0 \quad or \quad if \quad k = n \end{cases}$$

   Then $\hat{m}_k = m_k$ can be obtained when $\hat{\lambda}(k) = \hat{f}(t_k) - \hat{f}(t_{k-1}) > 0$. $\square$

The relationship between observations and the solution to an isotonic regression problem we found can be used to find the solution. We summarize it in the next theorem.

Theorem 3. *Assume T is a (finite) set of real numbers equipped with the simple order and $\mathcal{F}$ is the set of isotonic functions on T and y is an arbitrary functions on T. Then the solution to the isotonic regression problem*

$$\hat{f} = argmin_{\mathcal{F}} \sum_{t_i \in T} (y_i - z(t_i))^2$$

is obtained as the (left hand) slope of the greatest convex minorant of the cumulative sum diagram (k, Y_k). □

The geometric interpretation obtained can be used to drive three algorithms for isotonic regression: *Pool Adjacent Violators Algorithm* (PAVA), *Minimum Lower Set Algorithm* (MLSA), and *minmax* formulas.
The *Minimum Lower Set Algorithm* algorithm must start at the left of the data with the first point t_1. MLSA can be described as:

1. First start with the left point t_1, then add points until finding the position t'_1 that makes the average over $t_1, \ldots, t'_1$ as small as possible

$$t'_1 = \text{argmin}_{t_k} \frac{\sum_{i=1}^{k} t_i}{k}.$$

   This corresponds to making the slope from t_1 to t'_1 in the cumulative sum diagram as small as possible.

2. Next step is similar but starts with t'_1. Minimize the average over the lower sets $(t'1, \ldots, t_k)$ for $t_k \geq t'_1$ so

$$t'_2 = \text{argmin}_{t_k} \frac{\sum_{i=t'_1}^{k} t_i}{t'_2 - t'_1 + 1},$$

   this corresponds to making the slope from t'_1 to t'_2 as small as possible.

3. Go on through the whole sequence $t_1, \ldots, t_n$. Then the solution is obtained as a combination of piece-wised averages of original data.

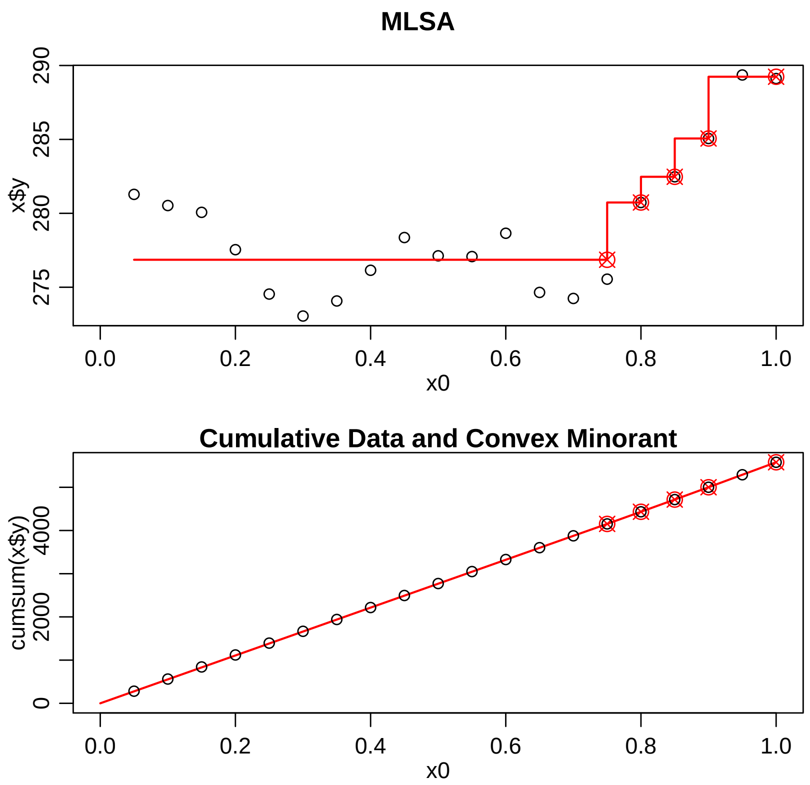

Figure 7: MLSA method

The PAVA algorithm can start from any position but is only valid for a simple order. First, if $t_i \leq t_j$, we define partial average

$$Av(t_i, \ldots, t_j) = \frac{\sum_{l=i}^{j} y_l}{j - i + 1}$$

Then the PAVA algorithm can be described as:

1. Calculate $Av(t_i, t_i)$ for each point. If the averages are ordered

$$Av(t_1, t_1) \leq \ldots \leq Av(t_n, t_n),$$

then y itself is an increasing function, and thus $\hat{f} = y$ is the solution to the isotonic regression problem.

2. If not, then for (at least) one k we have $Av(t_k, t_k) > An(t_{k+1}, t_{k+1})$. y is not increasing between t_k and t_{k+1}. The slop of cumulative sum diagram is decreasing. To make sure that the result line is below observation cumulative sum, replace both two averages with $Av(t_k, t_{k+1})$.

3. If the new averages are ordered, stop. If not, repeat step 2 until all the averages are ordered.

The solution to the isotonic problem is piecewise constant on the small intervals respectively (noting that theses intervals might be one-point sets).

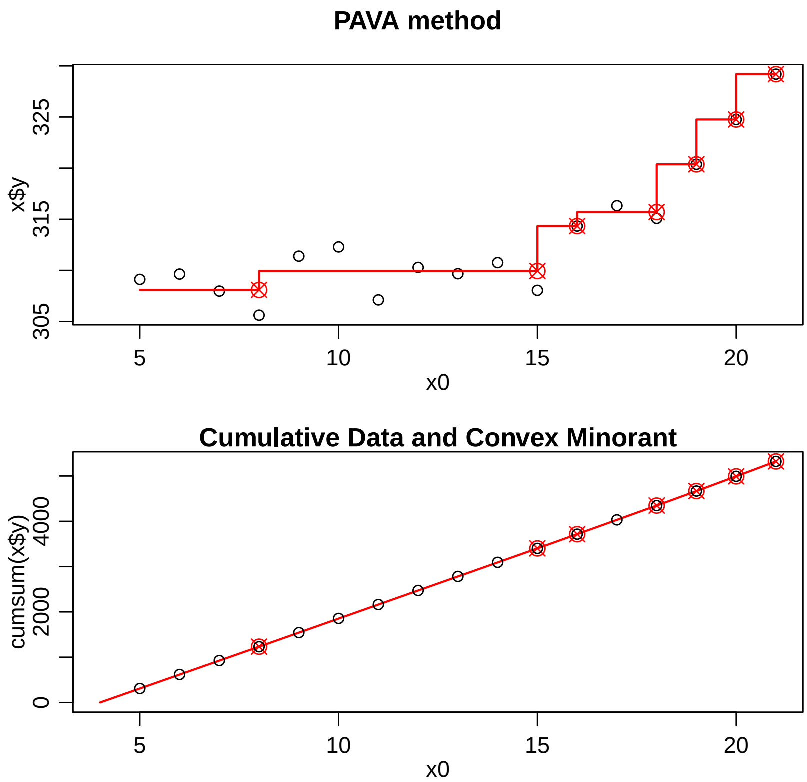

Figure 8: PAVA method

If we only want to calculate the solution at a fixed point t, and do not need the whole function, we can use the third algorithm : *minmax* formulas. They follow easily by studying the graph of the greatest convex minorant of the cumulative sum diagram, and are given by

$$\hat{f} = \begin{cases} max_{i \le k} min_{j \ge k} Av(t_i, \ldots, t_j), \\ max_{i \le k} min_{j \ge i} Av(t_i, \ldots, t_j), \\ max_{j \le k} min_{i \le k} Av(t_i, \ldots, t_j), \\ max_{j \le k} min_{i \ge j} Av(t_i, \ldots, t_j). \end{cases}$$

3.3 Taut string method

The taut strings method is an efficient way to fit an isotonic function. According to [3], the greatest convex minorant of the integrated data is a taut string and its derivative is precisely the least squares isotone approximation. The taut string method introduced here is based with assumptions in [7], you can find more details and proofs in [7]. In [11] , they first extended its use to the nonparametric regression.
Recall our problem is to find a function $f(t_i)$, such that data $y(t_i)$ could be written as

$$y_i = y(t_i) = f(t_i) + r(t_i)$$

which is a special case of the general Tukey decomposition

$$Data = Signal + Noise$$

Taut strings is based on Gaussian white noise. We stop algorithm when $r(t_i)$ may be adequately approximated by Gaussian white noise. Suppose $i = (1, 2, ..., n)$. The taut string method is to produce candidate functions f^k with k local extreme values and then to take the smallest K for which the residuals

$$r_i^k = y_i - f^k(t_i)$$

approximate white noise. Define the integrated process y_m° as

$$y_m^\circ = \frac{1}{n}\sum_{i=1}^{m} y_i, \qquad m = 1, \ldots, n.$$

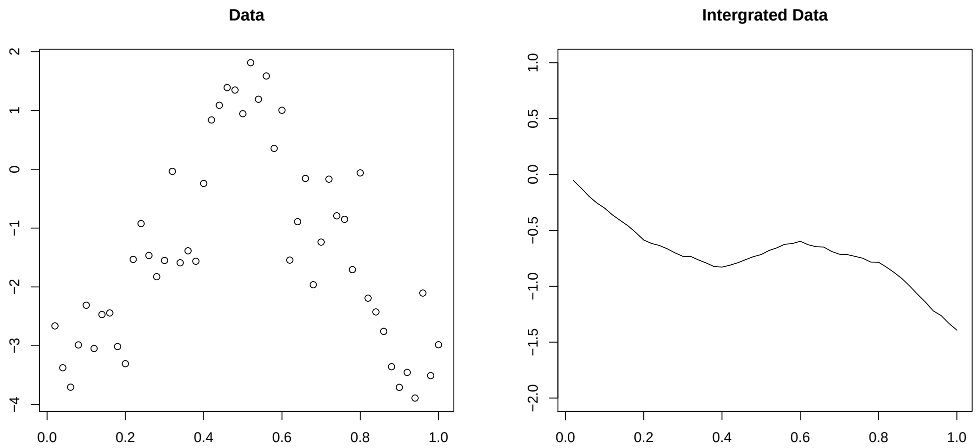

Figure 9: The left shows a small data set and the right panel shows their partial sums

Consider the lower bound l_n and upper bound u_n for y_m° defined by

$$l_n = y_m^\circ - \frac{C}{\sqrt{n}},$$

$$u_n = y_m^\circ + \frac{C}{\sqrt{n}}$$

for some $C > 0$. Image put a string between l_n and u_n, whose one end at the start point and the other at the end. Pull the string s until it is taut. The derivative of the taut string $\hat{f}^k = s'$ is an estimator to the data.

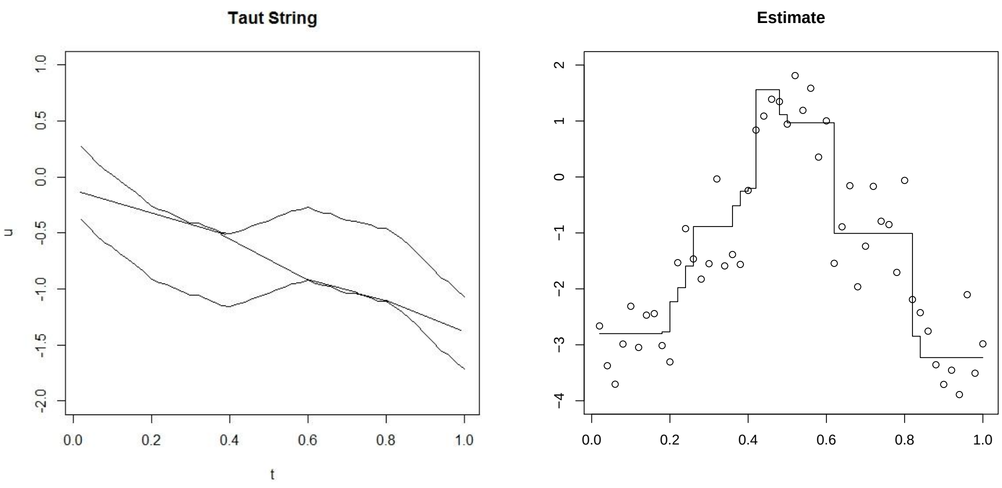

Figure 10: Taut string and the derivative of the taut string

The points at which the string coincides with either the lower and upper boundary are called knots. The function between knots is linear. The largest convex monorant of the upper bound is achieved between two knots where the string touches the upper bound and between which it does not touch the lower bound. Similarly, between knots where string touch lower bounds but not touch upper bounds, string is the smallest concave majorant of the lower bound. At points where the string switches from the upper bound to the lower bound the derivative s has a local maximum. Similarly, at points where the string switches from the lower bound to the upper bound the derivative s has a local minimum.
We need to consider local squeezing in taut string method, because a global radius for the tube will cause spurious local extremes. To solve this problem, we can accomplish local squeezing on the interval I by shifting the integrated process y_m° and then using the modified lower and upper bounds

$$l_n(t, I) = y_m^\circ(t) - \gamma,$$

$$u_n(t, I) = y_m^\circ(t) + \gamma$$

for $t \in I$. If the global tube radius is γ_0, which is $\frac{C}{\sqrt{n}}$, then we can define a local tube radius as $\gamma = \rho\gamma_0$. The research in [7] shows that, if ρ is around $\rho = 0.5$, then the procedure leads to too many local extreme values although cost less time. If ρ nearly 1, say $\rho = 0.95$, then the calculations has better estimation of local extreme values but requires more time.

3.4 Bootstrap

It is important to give an indication of the precision of a given estimate in most estimation problem [17]. An effective way to solve this problem is bootstrap. The bootstrap technique is a very general method to create measures of uncertainty and bias. Bradley Efron first introduce this new idea in the *Annals of Mathematical Statistics*, 1979. The word bootstrap hints at the saying "pull oneself up by ones bootstraps" [9].

A standard estimation problem can be describe as following: say we have some observations of distribution $F(x)$. Let $\hat{\theta}$ be an estimator of some parameter $\theta = G(F(\cdot))$. The distribution of the difference $\hat{\theta} - \theta$ contains all the information needed for assessing the precision of $\hat{\theta}$. For instance, we can construct a confidence interval of level $1 - \beta - \alpha$ as $[\hat{\theta} - \varepsilon_\beta \hat{\sigma}, \hat{\theta} - \varepsilon_{1-\alpha}\hat{\sigma}]$, where ε_β is the upper β-quantile of the distribution of $(\hat{\theta} - \theta)/\hat{\sigma}$. Here $\hat{\sigma}$ may be arbitrary, but it is typically an estimate of the standard deviation of $\hat{\theta}$. Unfortunately, in most cases the distribution $F(x)$ is unknown, so the quantiles and the distribution of $\hat{\theta} - \theta$ depended on $F(x)$ are unknown and cannot be used to assess the performance of $\hat{\theta}$.

The bootstrap technique replace the quantiles and distributions by estimators. The distribution of $(\hat{\theta} - \theta)/\hat{\sigma}$ under $F(x)$ can be written as a function of $F(x)$. The bootstrap estimator for this distribution is the "plug-in" estimator obtained by substituting $\hat{F}(x)$ for $F(x)$ in this function. Then we can obtain bootstrap estimator for quantiles and confidence intervals from the bootstrap estimator for the distribution.

Residual bootstrap is very useful in regression problem. Consider a linear regression problem as a simple example. There are n observations (x_i, y_i), where $i = 1, \ldots, n$. We want to estimate $y_i = \alpha + \beta x_i + \epsilon_i$ and we get estimator as $\hat{\alpha}$, $\hat{\beta}$ and $\hat{y}(x)$. So there are residuals after regression $e_i = y_i - \hat{y}_i(x)$. By doing N times bootstrap sampling of residuals, we will get N series of residuals as $e_1^{\star j}, \ldots, e_n^{\star j}$, where $j = 1, \ldots, N$. Using these series of residuals, we can construct N estimators of $\hat{y}(x)$ as $y^\star(x)^j$. There is a finding of the relation between these bootstrap estimator and original estimator of $y(x)$ as following:

$$n^{1/2}(\hat{y}(x) - y(x)) \longrightarrow Z \tag{8}$$

$$n^{1/2}(y^\star(x) - \hat{y}(x)) \longrightarrow Z \tag{9}$$

And Z has a Gaussian distribution. So we can do quantiles and confidence interval estimation for $y(x)$ based on the information obtained by $y^\star(x) - \hat{y}(x)$. Similar residual bootstrap can be used in our problem. Assume we have observations (y_i, t_i) and a estimator $\hat{f}(t)$. We will have residuals $e_i = y_i - \hat{f}(t_i)$, where $i = 1, \ldots, n$. After N times bootstrap sampling of residuals, we will get N series of residuals as $e_1^{\star j}, \ldots, e_n^{\star j}$, where $j = 1, \ldots, N$. Using these series of residuals, we can construct N estimators as $f^\star(t_i)^j$. Just like equations 8 and 9, we know a relation between these estimator in our case.

$$n^{1/3}(\hat{f}(t) - f(t)) \longrightarrow Z$$

$$n^{1/3}(f^\star(t) - \hat{t}(t)) \longrightarrow Z$$

where Z is depended on different situations. If t is away from a mode, then

1. if f is decreasing at t, $Z \sim T_{lcm}(-s^2 + W(s))'(0)$, where $W(s)$ is a a standard two-sided Brownian motion and "lcm" stands for least concave mironant. [19]

2. if f is increasing at t, $Z \sim T_{gcm}(s^2 + W(s))'(0)$, where $W(s)$ is a a standard two-sided Brownian motion and "gcm" stands for great convex monorant [1].

When t is at a mode, it will be a very complicated situation. Since we will not consider this case here, we will not explain more about this situation, but the interested reader can read reference [2] for this situation.
When we have these bootstrap estimator $f^\star$, we can do analysis about interested features in the result. For instance, we are interested in the value of the sixth peak y_p. We get $\hat{y}_p$ from taut string method and $y_p^\star$ can be obtained from $f^\star$. We can estimate the empirical distribution function and density function of $y_p^\star$ and construct a confidence interval of $\hat{y}_p$ using method above.

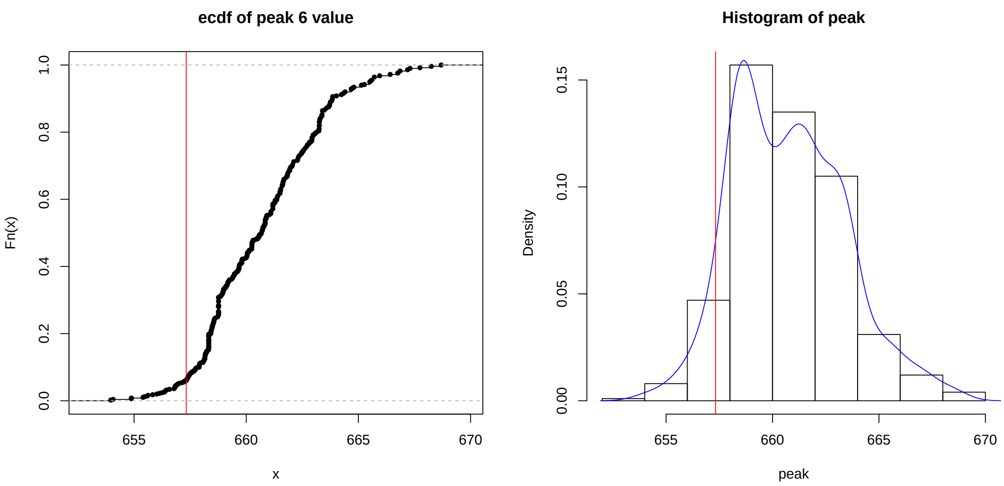

Figure 11: Left panel shows estimated ECDF and right panel shows estimated density function. The red line presents result from taut string method

4 Data Analysis and Results

To apply isotonic regression to the auditory responses data, I choose the PAVA algorithm. In order to decide connecting points at which two isotonic regression functions connect, first using kernel smoother to find these points. The drawback of this method in practice is that it needs much time and hard to judge if you choose right connecting points or not, since there is no strict condition about peaks and troughs.

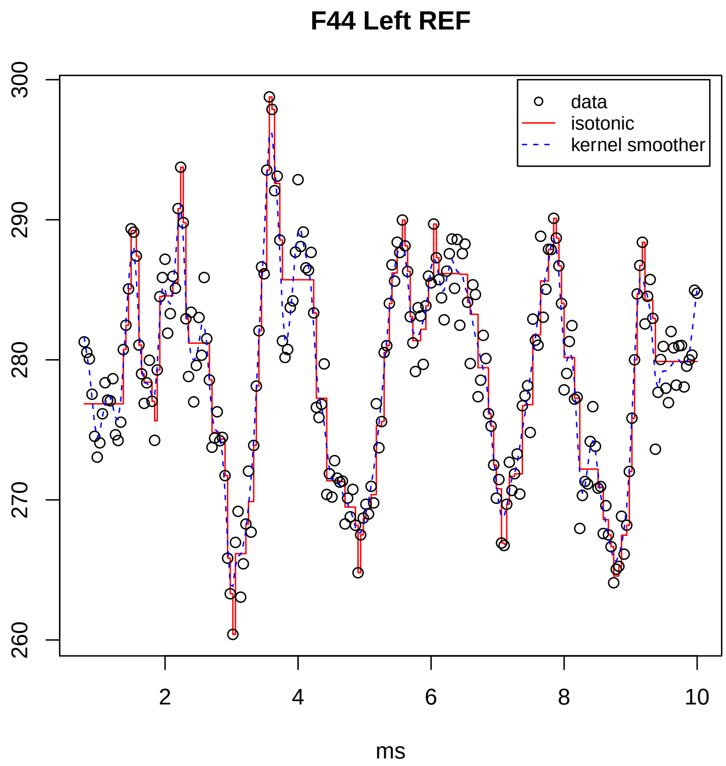

Figure 12: Isotonic regression

The taut string method shows advantages in this problem. It cost less time and has a similar performance to the isotonic regression. In the figure 13, you can see there is little difference between results of two methods.

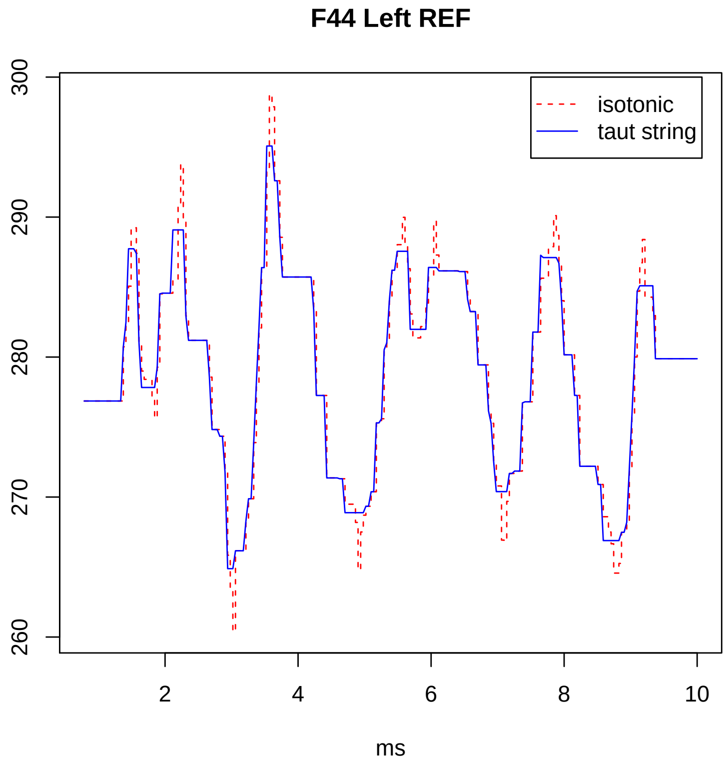

Figure 13: isotonic regression method and taut string method

According to the convenience of taut string method, we use results from taut string method to analysis in following parts.

4.1 Two peaks in first half

One feature of the result which we are interested in is if there is two obvious peaks in the first half, which means between 1 and 5 ms. From the results, we could find these peaks in patients and references in most pictures, but we failed in a female patient born in 1942 and a healthy male person born in 1940.

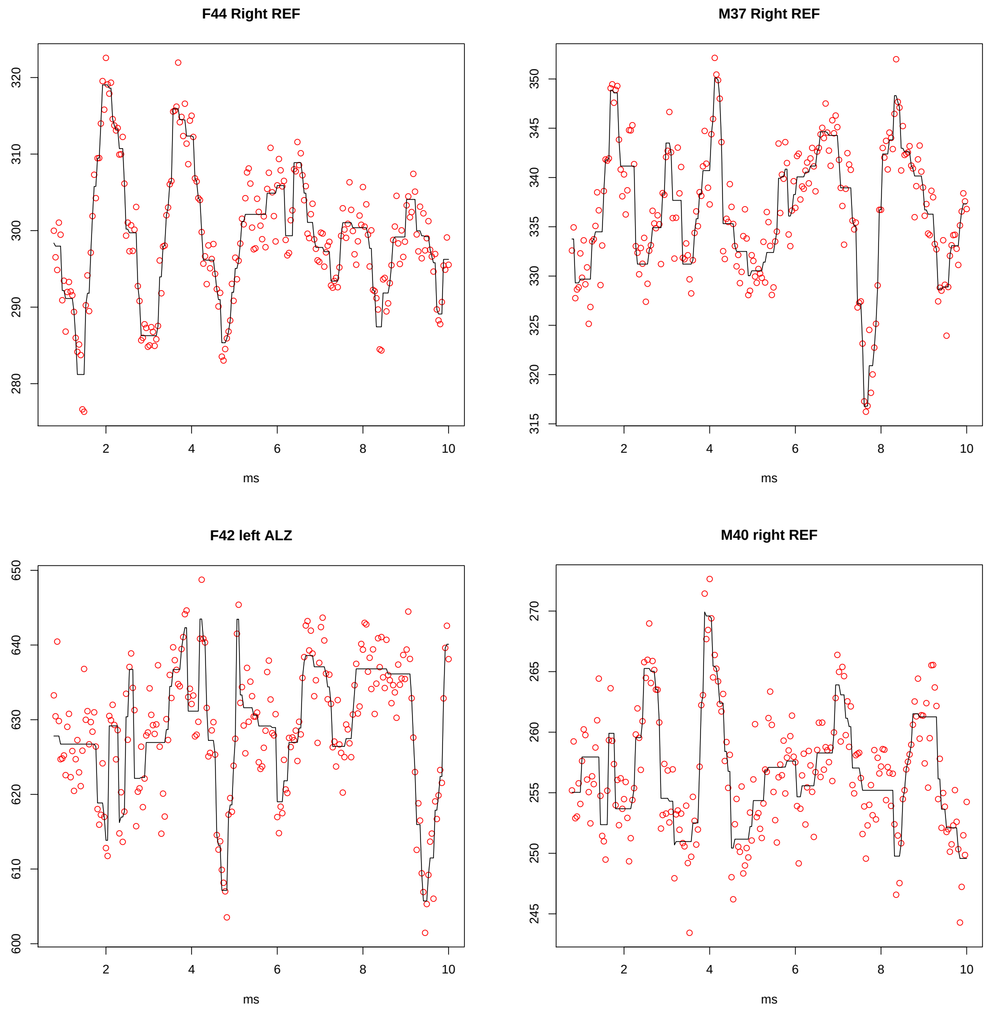

Figure 14: The upper two panels show obvious two peaks between 1 and 5 ms; the two panels below show 3 or more peaks

So this feature in not a good point to separate Alzheimer's patient from healthy people. If we want to use bootstrap technique here, a hypothesis test can be constructed as:

1. H_0: $z = 0$
2. H_1: $z \neq 0$

where

$$z = 1\{twopeaksfoundinreferencegroup\} - 1\{twopeaksfoundinpatientgroup\}$$

The difficulty here is that how to define a value is a local extreme, since it is necessary for constructing a automatic judging program for finding peaks. The value of peaks varies sharply among patients and healthy controls. This problem remains in next subsection also.

4.2 Two peaks around 7 ms

There are approximately six peaks in 10 millisecond period, and we could find two peaks around 7 ms in every picture. In some situation, it is easier to find two peaks that are closed, in both the patient and reference groups.

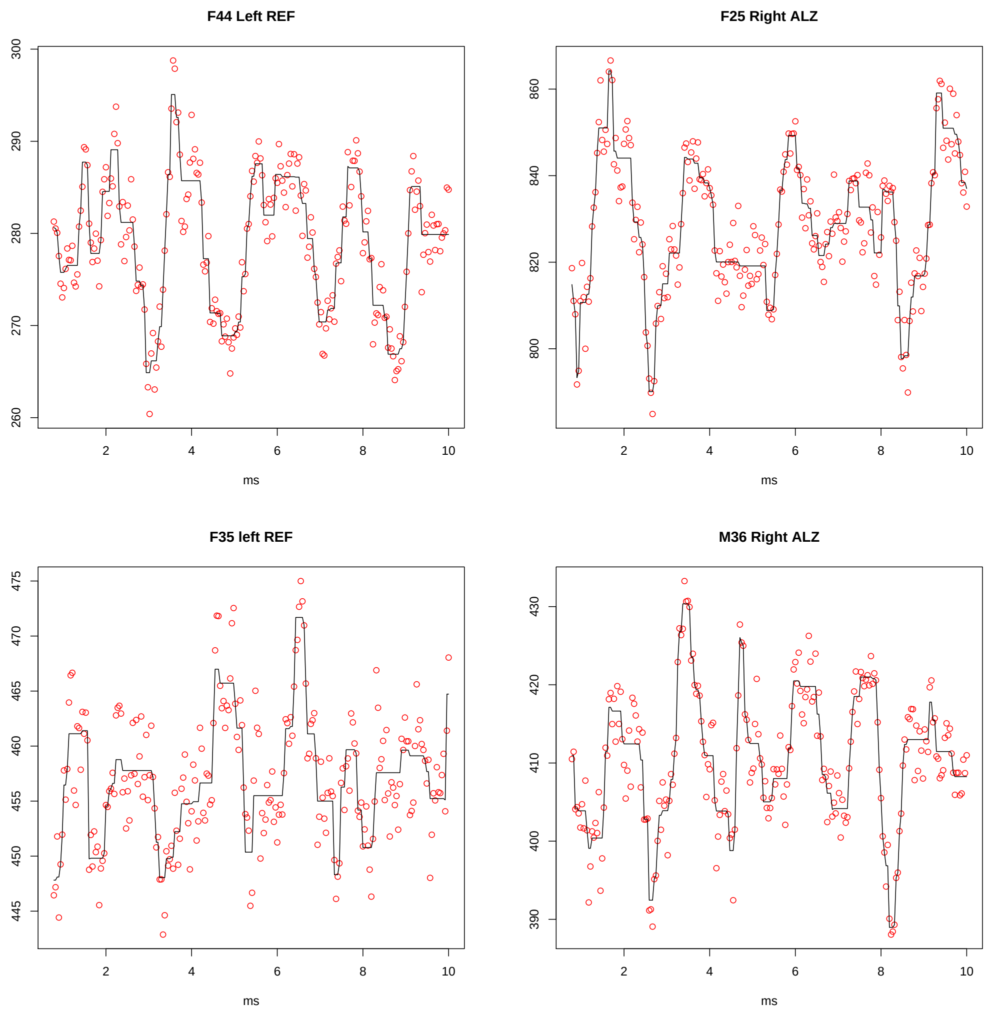

Figure 15: The upper two figures show obvious two peaks between 1 and 5 ms; the two pictures below show 3 or more peaks

4.3 The position of the trough between 4 and 6 ms

If we focus on time period between 4 and 6 millisecond, we will always find a trough after the third peak. In the reference group, the position of this trough is near 4.5 ms in most cases. While in the patient group, this position is around 4 ms in some cases. Sometimes, you even could not find a obvious trough in the result of a Alzheimer's patient.

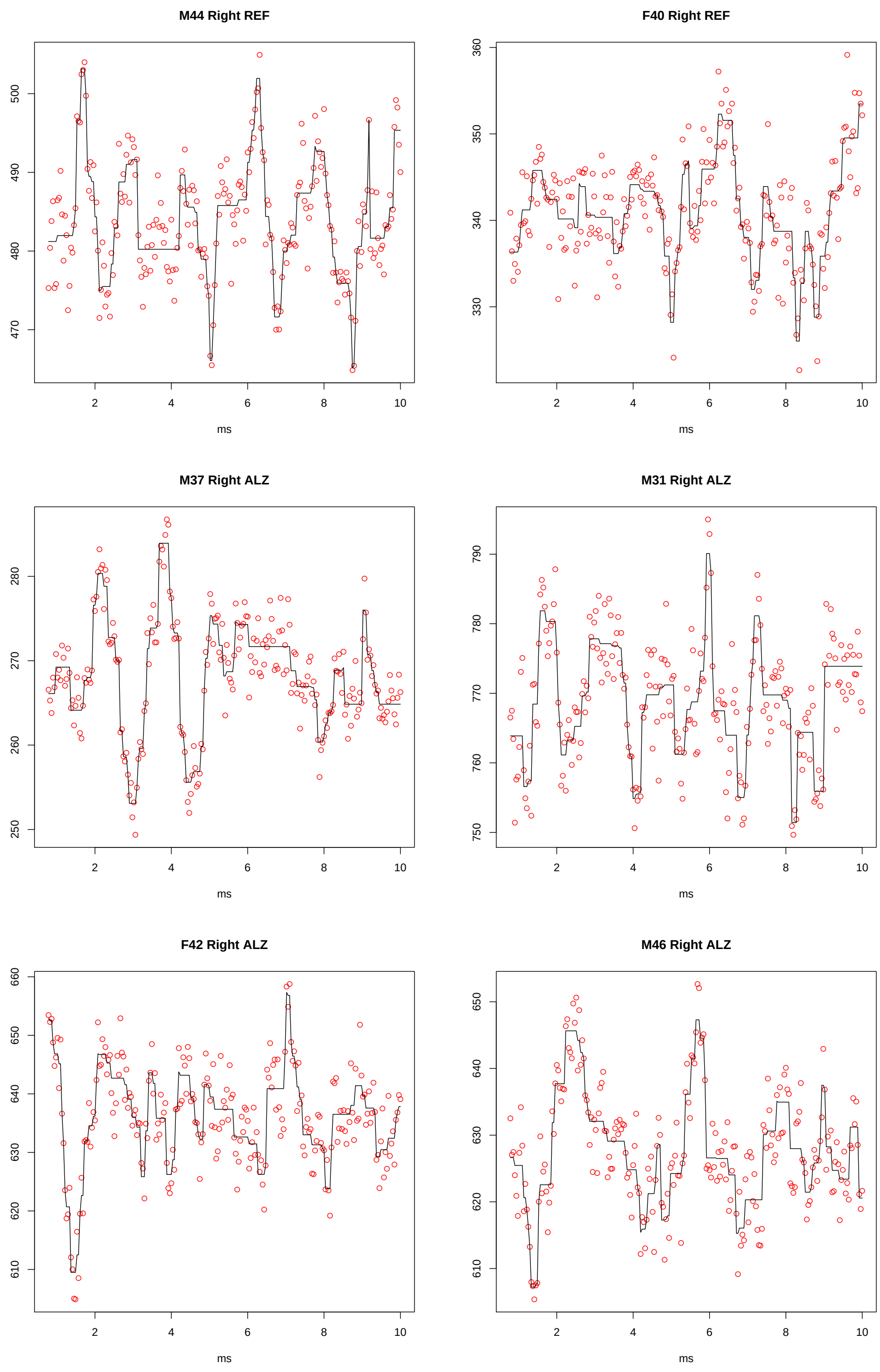

Figure 16: The upper two pictures show that the position of the trough is around 4.5 ms in health group; the two pictures in second row show that the position of the trough is near 4 ms in patient group; the two pictures below show that there is no obvious trough between 4 and 6 ms in some patients' results

If we want to do hypothesis test about if the position of the trough is the same between patients and healthy control, we could use help from bootstrap technique. For an example, we want to know the difference between a male patient born in 1937 and a healthy person in reference group born in the same year. Let $t = P_{ref} - P_{alz}$, where P_{ref} stands for the position of the trough from healthy control data and P_{alz} is the position from the Alzheimer patient. Our test is:

1. H_0: $t = 0$
2. H_1: $t \neq 0$

Set $N = 500$,and do residual bootstrap to the data. We can estimate empirical distribution function, density function of the bootstrap difference. The results are shown in figure 17.

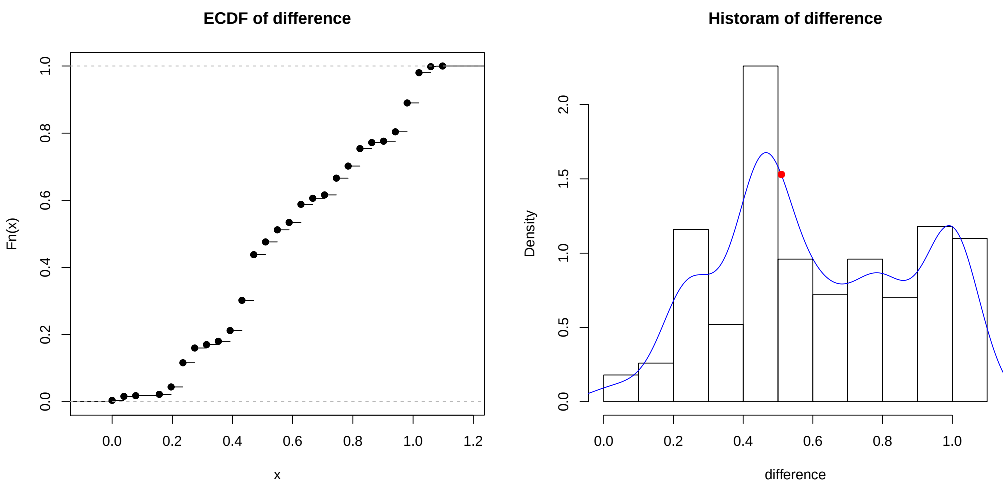

Figure 17: Left panel shows estimated ECDF and right panel shows estimated density function. The red point presents result from taut string method

From Ecdf we get, $p(H_0) = 0$. So reject H_0, which means there is significant different position of the trough in two groups.
From the taut string results, the difference between two people is $\hat{t} = 0.51$. After residual bootstrap, a confidence interval of t can be accomplished as $[\hat{t} - \varepsilon_{0.025} sd(t^\star), \hat{t} + \varepsilon_{0.975} sd(t^\star)]$, where ε_β is the upper β-quantile of the distribution of $(t^\star - \hat{t})/sd(t^\star)$. So we can get confidence interval as $[0.31, 1.52]$.

4.4 Range of y-axis

Another efficient way to separate Alzheimer's patient from healthy people is checking the range of y-axis. For healthy people, the start value of the y-axis lies between 260 to 480, usually between 280 to 330. While in patient group, the value could start 600, even reach 800. As to the range of y-axis, in healthy group, the range is between 30 to 40. While in patient group, the range increases to 60 or 70 sometimes.

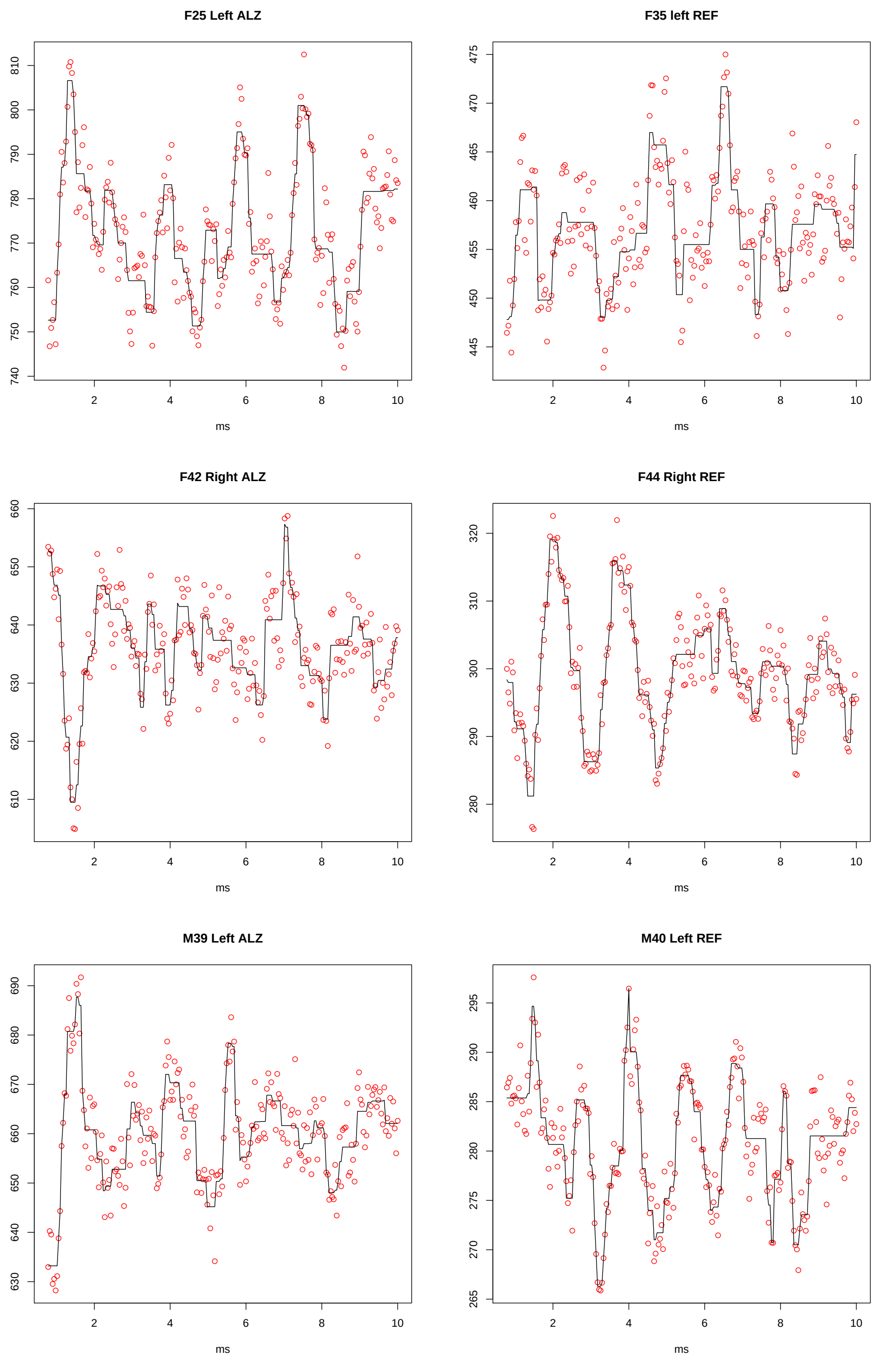

Figure 18: The left panels show results of Alzheimer's patients and the right panels show results of reference group

4.5 Difference between left and right ears

A suggestion of the construction of the brain is like the picture below:

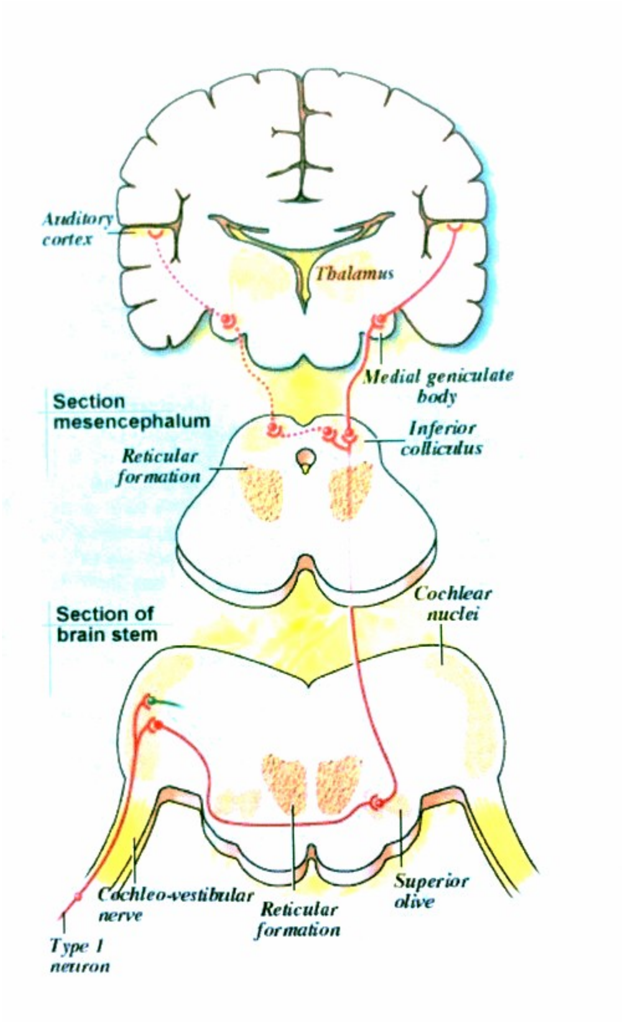

Figure 19: brain construction

We could check if signals pass same area no matter which side it starts by comparing two pictures of different ears of one same person. The results in figure 20 show that, area which third and fourth peak happens is shared by both side signal.

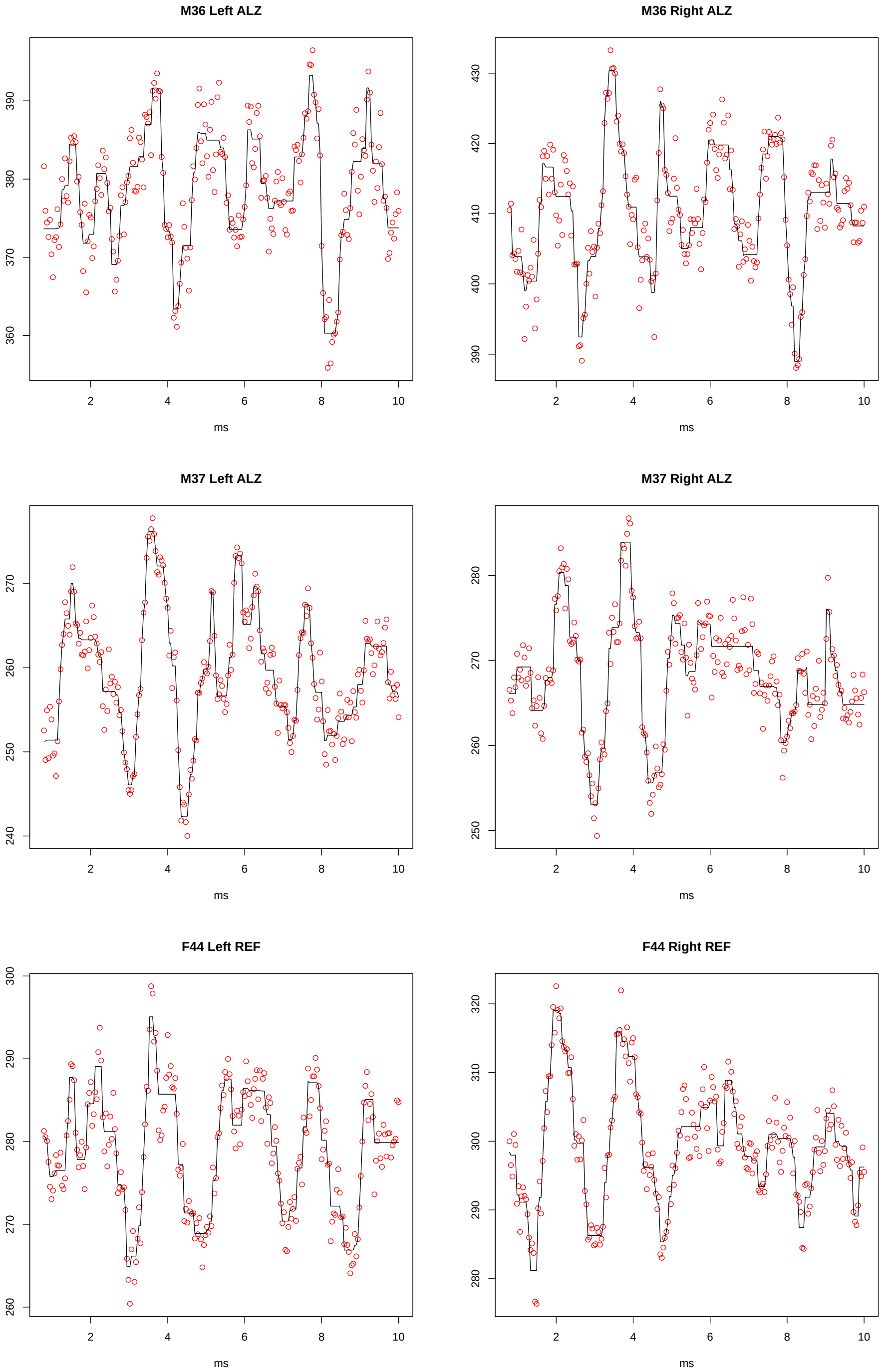

Figure 20: The left panels show results of left ears data and the right panels show results of right ears data

5 Conclusion

Relying on advantages that not easily affected by general anesthetics, the level of consciousness nor the cooperation of the subject, brainstem auditory responses supply a new and promising direction for researches in Alzheimer disease. According to the steady result of each individual, the features which are found in the results are stationary and reliable.
Isotonic regression and taut strings are two useful methods to model brainstem auditory responses. The solution to a isotonic regression problem is obtained as the slope of the greatest convex minorant of the cumulative sum diagram. On the other side, taut string method works as pulling a string until it is taut between upper and lower bound for the integrated data and the derivative of this string is the estimator that we are looking for. The two methods shares similar principle, but isotonic regression works as a local modeling while taut string method acts as a global modeling.
Local squeezing should be considered in this problem using taut string method. If we use a global tube radius γ_0 to decide the distance between upper and lower bound of the integrated data, spurious local extremes will appear. So we reduce it to a local tube radius as $\gamma = \rho\gamma_0$. It turns out that the calculations has better estimation of local extreme value and requires tolerable time when $\rho = 0.95$.
As shown in figure 21, we can find just a little difference in the results of two methods except the points where extreme values occur. The reason is that isotonic regression method fits a model which reaching the peaks and troughs of the original data. But taut string method use a tube radius to decide the distance between upper and lower bound of the integrated data and calculate the derivative then, which cannot reach the extreme values of the original data. Even though taut string method presents disadvantage on extreme values, according to costing less time in practical use, we choose taut string method in the situation when the positions of extreme values are important and numerical values of extreme values are not considered.

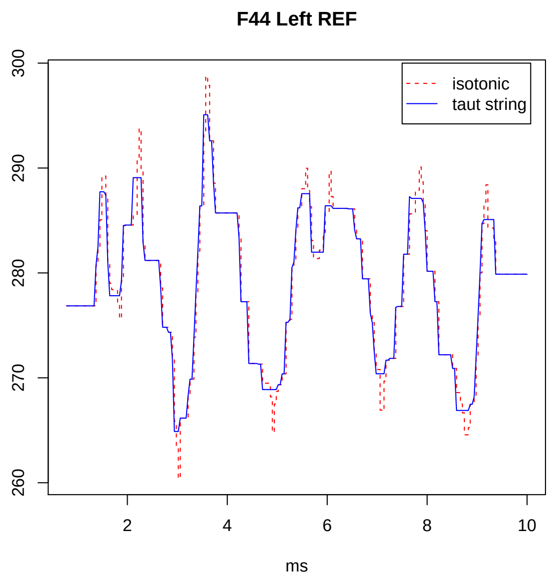

Figure 21: isotonic regression method and taut string method

Bootstrap technique helps to analyze results in this problem. Residual bootstrap provides $f^\star$ which are helpful to estimate empirical distribution function, density function and confidence interval of our taut string solution $\hat{f}$. Thanks to bootstrap technique, we have much clearer results in feature analysis. But there are some problem remained to be solved about the construction of the test statistics.

Final results show that there should be two brainstem waves between 1 and 5 ms and another two around 7 ms in any individual record. Two obvious differences between Alzheimer's patients and healthy people are, the time at which the third trough happened and the y-axis range. We can also check that no matter which ear it starts from, signals pass the same area in brainstem in the middle of whole process. We can find more interested features once we have a nonparametric model $\hat{f}$ to describe data. Sometimes we can reduce the complexity of $\hat{f}$ to just a few numbers of features and sometimes the features we are looking for can even be functions of the solution $\hat{f}$ itself.

References

[1] D. Anevski and O. Hössjer. A general asymptotic scheme for inference under order restrictions. *Ann. Statist.*, 34(4):1874–1930, 2006.

[2] Fadoua Balabdaoui, Hanna Jankowski, Marios Pavlides, Arseni Seregin, and Jon Wellner. On the Grenander estimator at zero. *Statist. Sinica*, 21(2):873–899, 2011.

[3] R. E. Barlow, D. J. Bartholomew, J. M. Bremner, and H. D. Brunk. *Statistical inference under order restrictions. The theory and application of isotonic regression.* John Wiley & Sons, London-New York-Sydney, 1972. Wiley Series in Probability and Mathematical Statistics.

[4] Johnson E Brookmeyer R and MH Arrighi Ziegler-Graham K. Forecasting the global burden of alzheimer's disease. *Alzheimer's and Dementia*, 3(3):186–191, July 2007.

[5] Kawas C. Brookmeyer R., Gray S. Projections of alzheimer's disease in the united states and the public health impact of delaying disease onset. *American Journal of Public Health*, 88:1337–1342, 1998.

[6] Keith H. Chiappa. *Evoked Potentials in Clinical Medicine.* Lippincott-Raven Publishers, third edition, 1997.

[7] P. L. Davies and A. Kovac. Local extremes, runs, strings and multiresolution. *Ann. Statist.*, 29:1–65, 2001.

[8] Waldemar G. Recommendations for the diagnosis and management of alzheimer's disease and other disorders associated with dementia: Efns guideline. *Eur J Neurol*, 14:e1–26, January 2007.

[9] J. S. Urban Hjorth. *Computer intensive statistical methods.* Chapman & Hall, 1994. Validation model selection and bootstrap.

[10] Mia Ling Skold Johan Kallstrand, Sara Fristedt Nehlstedt and Soren Nielzen. Lateral asymmetry and reduced forward masking effect in early brainstem auditory evoked responses in schizophrenia. *Psychiatry Res.*, 2011.

[11] Enno Mammen and Sara van de Geer. Locally adaptive regression splines. *Ann. Statist.*, 25:387–413, 1997.

[12] Schumock GT Meek PD, McKeithan K. Economic considerations in alzheimer's disease. *Pharmacotherapy*, 18(2):68–73, Mar-Apr 1998.

[13] Rinne UK Mölsä PK, Marttila RJ. Survival and cause of death in alzheimer's disease and multi-infarct dementia. *Acta Neurol Scand*, 74:103–107, August 1986.

[14] Rinne UK Mölsä PK, Marttila RJ. Survival and cause of death in alzheimer's disease and multi-infarct dementia. *Acta Neurol Scand*, 91:159–164, March 1995.

[15] Tiraboschi P and Corey-Bloom J Hansen LA, Thal LJ. The importance of neuritic plaques and tangles to the development and evolution of ad. *Neurology*, 62(11):1984–1989, June 2004.

[16] Tim Robertson, F. T. Wright, and R. L. Dykstra. *Order restricted statistical inference*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Ltd., 1988.

[17] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 1998.

[18] Winblad B Wimo A, Jonsson L. An estimate of the worldwide prevalence and direct costs of dementia in 2003. *Dement Geriatr Cogn Disord*, 21(3):175–181, January 2006.

[19] F. T. Wright. The asymptotic behavior of monotone regression estimates. *Ann. Statist.*, 9(2):443–448, 1981.