Surface Representations of Sound

Mengyi Shan, Luke Trujillo

May 6, 2020

Surface Representations

In the paper we studied, the authors created created surfaces of spectrograms and implemented a time warping algorithm.

Given an audio signal $x(t)$, the authors first generated spectrograms by computing the **local Fourier transform**, given by

$$X(\omega, t) = \int_{-\infty}^{\infty} x(t)\psi(\tau - t)\exp(-j\omega\tau)d\tau$$

taken at frequency ω and time t using a Gaussian window function with size 10 milliseconds.

$$\psi(\tau) = \exp\left(-\frac{1}{2}\left(\frac{\tau}{0.5(10^{-0.2})}\right)^2\right).$$

Surface Representations

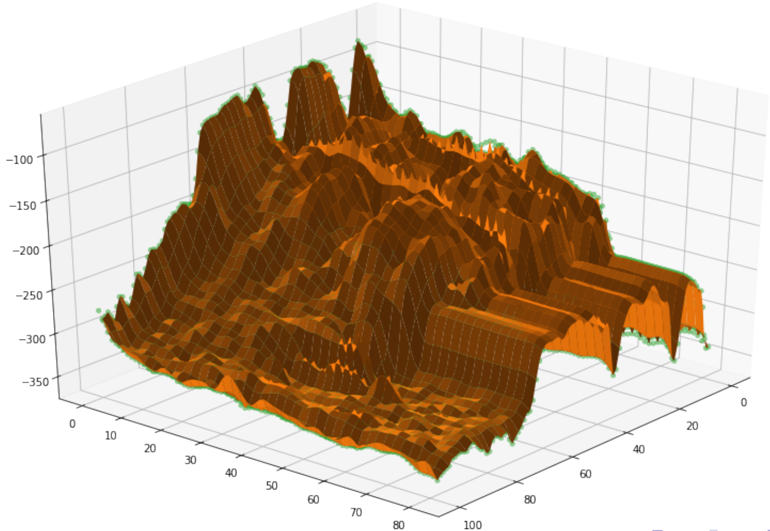
The authors then computed the **power spectral density** of the signal:

$$P(\omega, t) = 10 \log_{10}(|X(\omega, t)|^2)$$

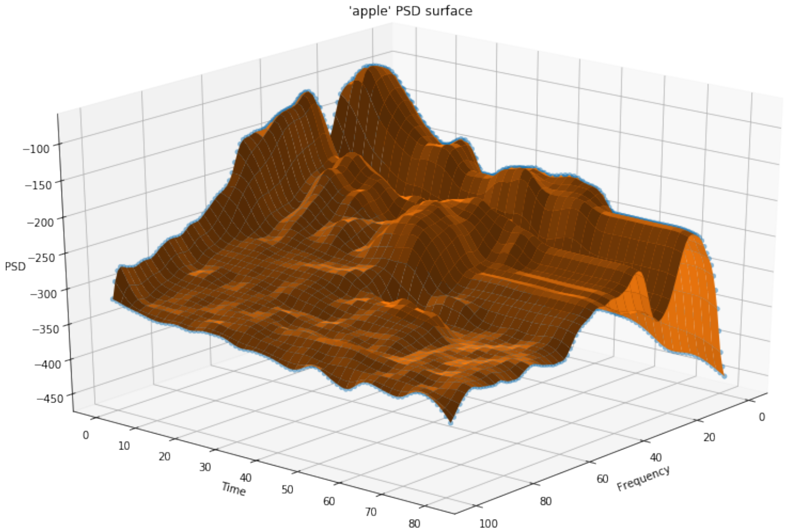
and created surface representations of sound by corresponding them to the height functions $(t, \omega, P(\omega, t))$.

Spectrograms themselves are three dimensional surfaces projected to $\mathbb{R}^2$. The authors are simply rescaling them in $\mathbb{R}^3$. However, this immediately leads to issues with noise and smoothing.

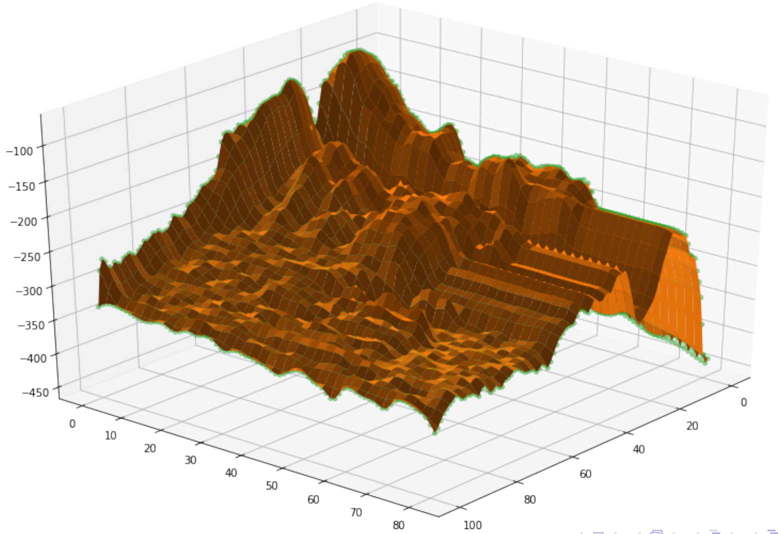
Smoothed PSD for "Potato"



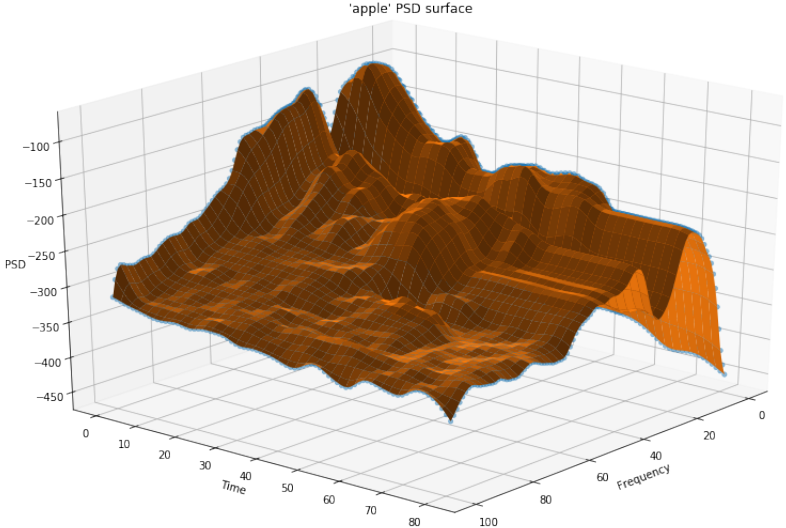
Smoothed PSD for "Potato"



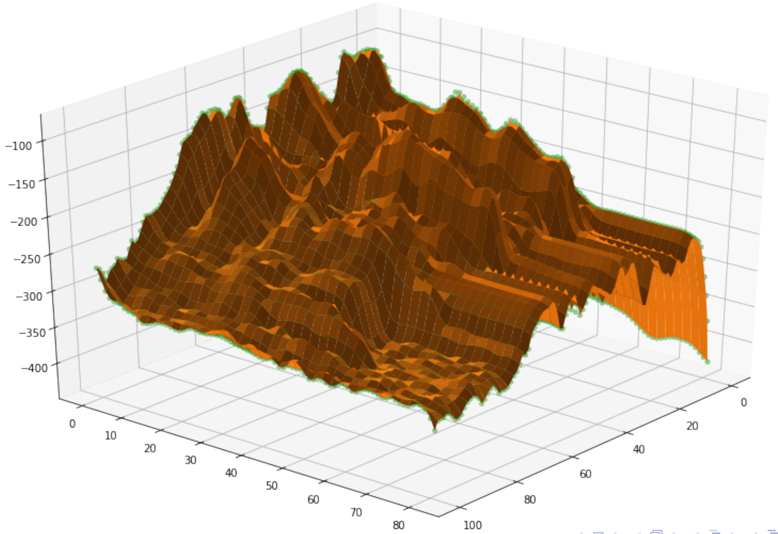
Smoothed PSD for "Apple"



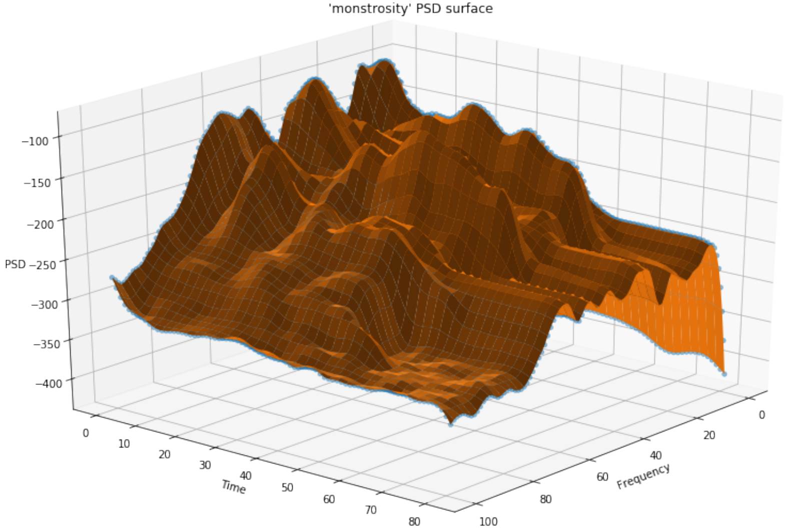
Smoothed PSD for "Apple"



Smoothed PSD for "Monstrosity"



Smoothed PSD for "Monstrosity"



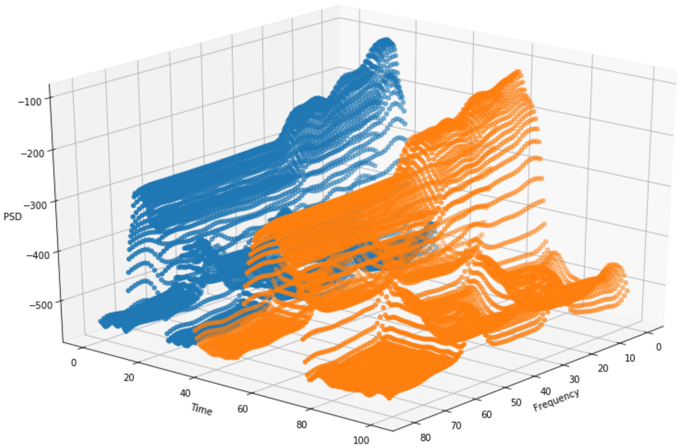
Concatenating Surfaces

Concatenating surfaces is a difficult challenge, but it is simplified if we only scale along the time axis which helps preserve the information.

Peaks of our PSD surfaces are almost always located along the lowest frequency axis. Therefore we scaled our surfaces by aliging their maxima on these axes, while trying to avoid scaling too much.

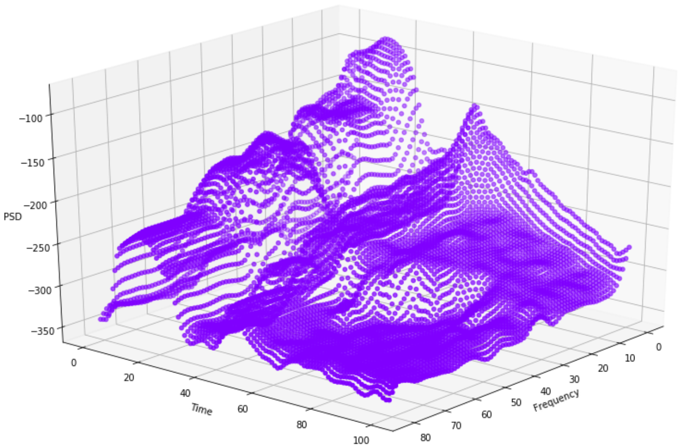
Concatenating Surfaces

The phonemes "u" and "p" concatenated together



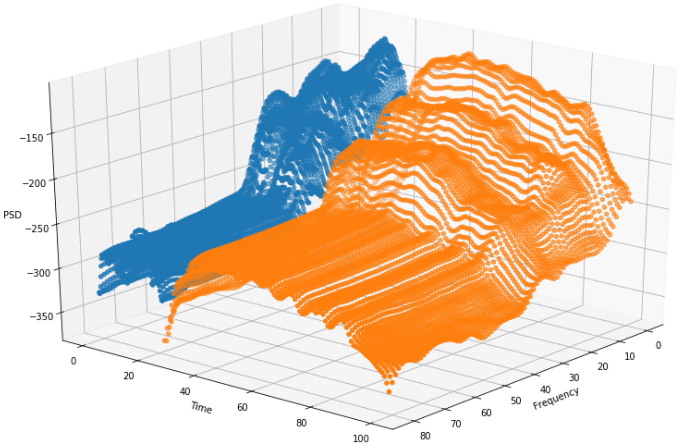
Concatenating Surfaces

The plot for the word "up", generated by phonemes "u" and "p".



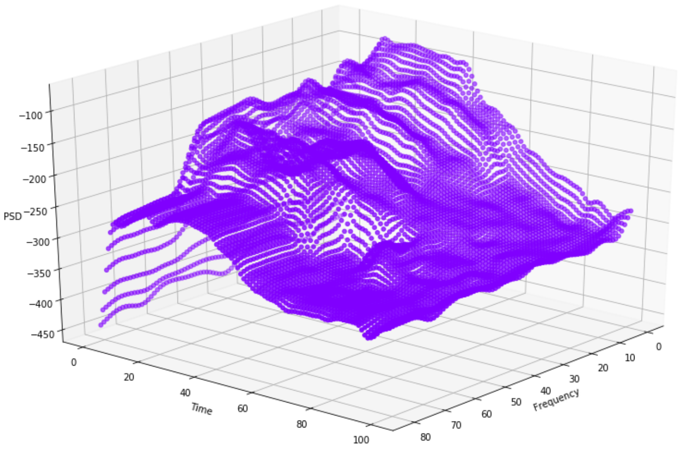
Concatenating Surfaces

Concatenation of the phonemes "k" and "a:" (sounds like "are")



Concatenating Surfaces

The plot of the word "car" made of phonemes "k" and "a:".



Persistence Homology

Let X be a topological space. Consider the n-th homology group

$$H_n(X)$$

and suppose that $f : X \to \mathbb{R}$ is a real-valued function. For any $a \in \mathbb{R}$, consider the topological space

$$f^{-1}((\infty, a]) \subset X$$

which inherits the subspace topology from X. Observe that if $a \leq b$ then this induces a function

$$i : f^{-1}((\infty, a]) \to f^{-1}((\infty, b])$$

namely, the inclusion function.

Persistence Homology

Now consider the homology group of this subspace

$$H_n(f^{-1}((\infty, a]))$$

and observe that for any $a \leq b$, we have a group homomorphism which we denote as φ_a^b:

$$\varphi_a^b : H(f^{-1}(\infty, a]) \to H(f^{-1}(\infty, b]).$$

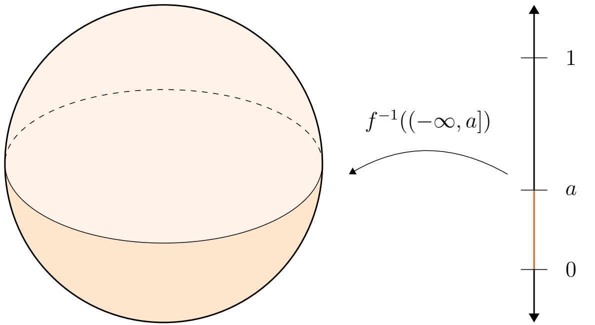
What we've created is a functorial data pipeline

$$a \longmapsto f^{-1}((\infty, a]) \longmapsto H(f^{-1}((\infty, a])).$$

which sends numbers to topological spaces to abelian groups. What this records is the evolution of the homology of our function defined on the topological space!

Persistence Homology

An example of this pipeline is if X is a sphere and $f : X \to \mathbb{R}$ is the height function.



As a increases, we can keep track of the homology groups to understand our data better.
But, for nice spaces (e.g. simplicial complexes), the homology usually won't change much.

Persistence Homology

What we mean is that, in general,

$$\varphi_a^b : H(f^{-1}(\infty, a]) \to H(f^{-1}(\infty, b]).$$

is usually an isomorphism. Therefore we define

$$\begin{aligned}\beta_a^b &= \operatorname{rank}(\operatorname{Im}(\varphi_a^b)) \\ &= \operatorname{rank}(\operatorname{Im}\Big(H(f^{-1}((\infty, a])) \to H(f^{-1}((\infty, b]))\Big))\end{aligned}$$

to be the **Betti number from** a **to** b and we pay attention to whether or not this number changes.

Persistence Homology

Moreover, if we do find an a such that, for some ε, the homomorphism

$$H_n(f^{-1}((\infty, a-\varepsilon])) \to H_n(f^{-1}((\infty, a+\varepsilon]))$$

is *not* an isomorphism, then we say a is a **critical value**. This means something happened to our homology; i.e., a singularity (like a hole or a vertex) was encountered.

Researchers were aware of this pipeline for sometime, but it was only in the 2000's that applied topologists created an extremely useful and stable tool for efficiently utilizing this technology.

Persistence Homology

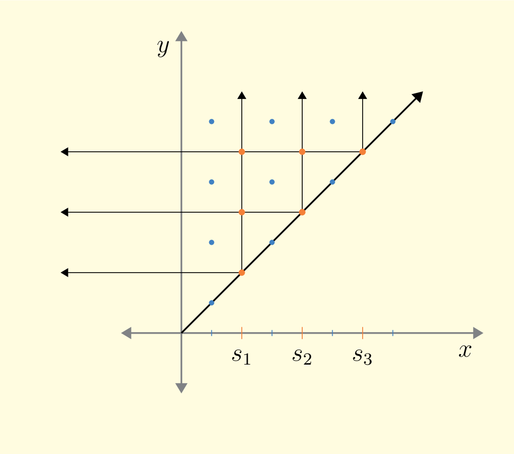
Suppose we have finitely many critical values $\{s_1, s_2, \ldots, s_n\}$. Let $\{t_0, t_1, \ldots, t_n\}$ be any interleaved sequence of numbers such that $t_{i-1} < s_i < t_i$.

Let $f : X \to \mathbb{R}$ have finitely many critical values and let (s_i, s_j) be a tuple of critical values. Then we define the **multiplicity** of (s_i, s_j) to be

$$\mu_i^j = \beta_{t_{i-1}}^{t_i} - \beta_{b_i}^{b_j} + \beta_{b_i}^{b_{j-1}} - \beta_{b_i}^{b_j}$$

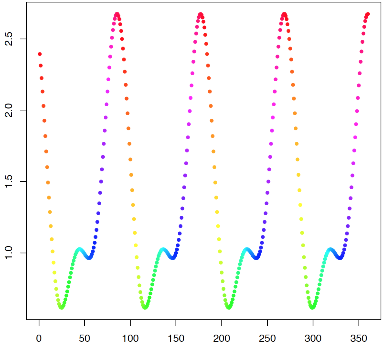
Persistence Homology

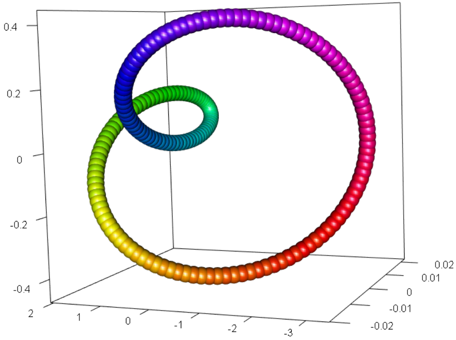
The persistence diagram of the tame function $f : X \to \mathbb{R}$, denoted $D(f)$, is the *multiset* of tuples (s_i, s_j) each with multiplicity μ_i^j.



Persistence Homology

In recent years, persistence homology has become an extremely useful tool for time-series analysis. Given a dataset, a sliding window embedding can be performed to create a point-cloud:

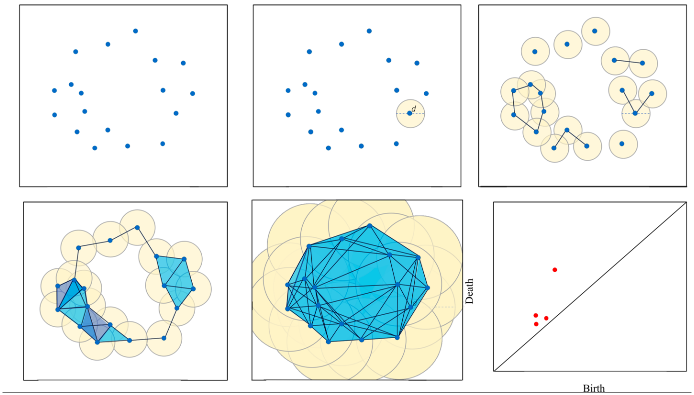




The homology of the resulting object is then studied.

Persistence Homology

In the case of real data, the homology of the data is filtered as a simplicial complex.



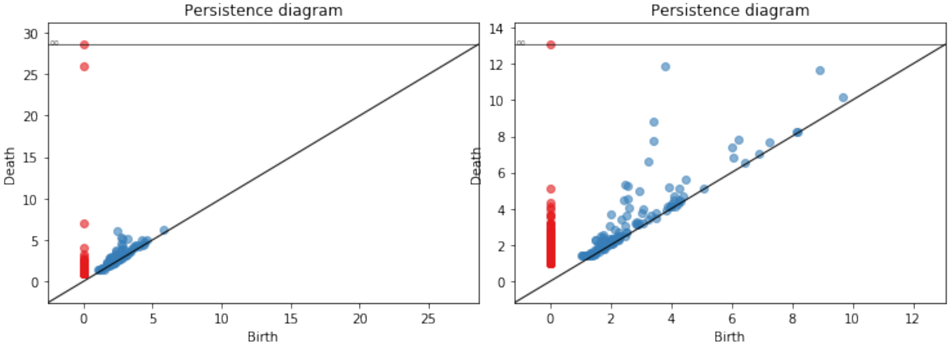
The most popular filtration is the Vietoris-Rips, which we used in our project.

Coding

- We utilized the GUDHI (Geometric Understanding in Higher Dimensions) python package for the project
- Used $\mathbb{Z}_2$ homology, which is fast and also intuitive
- We filted our data using the Vietoris Rips complex.

Vowels

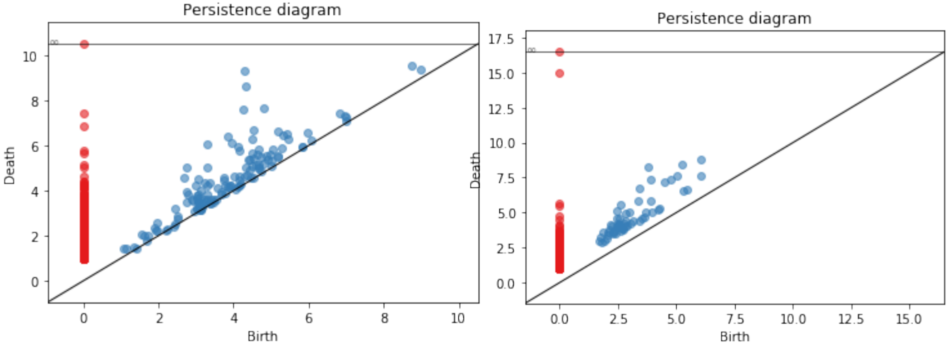
There is a significant difference in the persistence diagrams between vowels and consonants:



The above left is for the phoneme "k" while the one on the right is for the phoneme "a:" (sounds like the word "are").

Vowel Strength

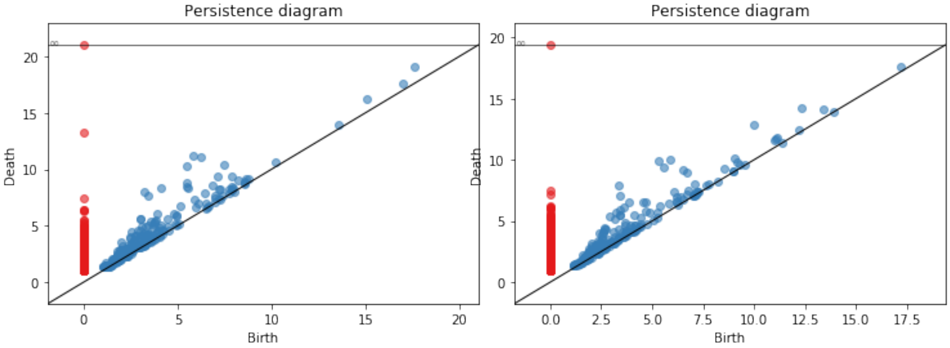
The difference is also apparent when considering words with vowels of different strengths.



The above left is the word "no", which has a strong vowel, while the one of the right is "yes", which is weaker.

Rhymes

We also checked words that rhyme to see if their persistence diagrams are similar.



The above left is the word "tarnation", while the above right is for the word "vacation."