UNIVERSIDADE DE SÃO PAULO
Faculdade de Zootecnia e Engenharia de Alimentos

RAFAEL ZINNI LOPES

Evaluation of sensory crispness of dry crispy foods by convolutional neural networks

PIRASSUNUNGA – SP
2023

Rafael Zinni Lopes

Evaluation of sensory crispness of dry crispy foods by convolutional neural networks

"Versão Corrigida"

Master thesis presented to the Faculty of Animal Science and Food Engineering of the University of São Paulo, as part of the requirements for obtaining the title of Master in Engineering and Materials Science.

Area of Concentration: Engineering and Materials Science

Head Advisor: Ph.D. Prof. Gustavo César Dacanal

PIRASSUNUNGA – SP

2023

Ficha catalográfica elaborada pelo
Serviço de Biblioteca e Informação, FZEA/USP,
com os dados fornecidos pelo(a) autor(a)

Lopes, Rafael Zinni
L864e Evaluation of sensory crispness of dry crispy foods by convolutional neural networks / Rafael Zinni Lopes ; orientador Gustavo César Dacanal. -- Pirassununga, 2023.
74 f.

Dissertação (Mestrado - Programa de Pós-Graduação em Engenharia e Ciência de Materiais) -- Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo.

1. Crocância. 2. Rede Neural Convolucional. 3. Librosa. 4. Pão. 5. Batata Frita. I. Dacanal, Gustavo César, orient. II. Título.

PROJETO DE DISSERTAÇÃO DE MESTRADO

Título: “Avaliação da crocância sensorial de alimentos crocantes secos por redes neurais convolucionais”

Agências de fomento: O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) - Código de Financiamento 001 e pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) - processo 2021/05317-9.

ACKNOWLEDGEMENTS

When people ask what is my purpose, I state: “I want to make everyone in the world to be capable of seeing, listening, and speaking.” You would think why I worked on a project that’s different from what I want to achieve soon. I thank you reader for thinking this way, we are here to learn that changing our perspectives brings us to our dreams. Life isn’t linear, neither dreams nor plans.

The results of this work are not mine alone, but of every person who gave an opinion. We may go faster alone, but together we can reach unexpected heights.

I thank Professor Gelson Andrade da Conceição for being more than a teacher, but a friend who guided my heart and mind toward what was best for me. It was hard to transform my fixed mindset into a growth mindset. You prepared me to overcome the most difficult barrier that was making me regress in life: myself.

Victor Dias de Oliveira, my best friend, you were always backing me up when I needed it. A few words can’t describe how grateful I am for having you by my side. You have built a marketable version of the crispness analyzer, it’s an honor to work with you to bring our initiatives into reality.

I thank Professor Gustavo César Dacanal for the opportunity to work on this dissertation which is opening new opportunities at LAFLUSP. This work complemented with praise the areas that I seek to develop as a professional: the developing of new products, neural networks, and programming logic.

Vivian Lara, you have planted the seed of entrepreneurship in my mind and my heart, I don’t doubt that this work one day will be a reference in the control quality laboratories around Brazil, just wait and you will see.

Daniela Correa and Larissa Rodrigues, you two have accompanied me in the laboratory days, that big yet hollow place, it’s because of you two that we could bring energy to that place.

I thank my family for supporting me.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and FAPESP (São Paulo Research Foundation) under grant 2021/05317-9, for their financial support. I’m grateful for their financial support.

I thank God for always bringing me challenges and life-learning opportunities, one day the talents you gave me will bring a huge positive impact on society.

“Seu maior diferencial competitivo é ser você. Só você pensa como você. Só você tem suas aptidões, habilidades e viveu as experiências que você viveu. Se permita ser quem você é, as pessoas que lutem para te aceitar ou te engolir do seu jeito.”

Davi Braga

ABSTRACT

LOPES, R. Z. **Evaluation of sensory crispness of dry crispy foods by convolutional neural networks**. 2023. 67 f. Master thesis – Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, 2023.

Convective drying is traditionally used to dehydrate food, reducing volume and water activity for easy transportation and storage. During drying, foods undergo volume reduction due to moisture loss, resulting in changes in the solid matrix and the formation of a crispy structure when crushed or fractured. This study focused on developing methods for quantifying and classifying crispy dried foods, such as potato chips, toasts, and fried foods like french fries and fried chicken, which were investigated. Compression profiles and sound noise were determined using a lever device covered by a noise suppression box. The captured sound was transformed into different parameters using Python and Mathematica Wolfram libraries. The power spectrum of the sound signal was obtained using the discrete Fourier transform method in Wolfram, while Onset Strength and Mel Frequency Cepstral Coefficients (MFCC) were obtained using the Librosa library. The sound spectra, Onset Strength, and MFCC were processed using neural networks to classify the crispness of fried chicken, potato chips, and toasts. The classification models using DFT and MFCC signals achieved an accuracy of over 95%. This study allowed the description of crispy sounds based on the intensity and duration of the signal. A second study utilized Python code and the Librosa library in an attempt to generate a dimensionless number, called the Zeta value, for classifying crispness intensity. The Zeta value was calculated based on Root Mean Squared Energy values multiplied by peak intensities within 1-second intervals. Experimental validation of the Zeta value was performed by acquiring crispness noises for toasts and French fries while monitoring moisture and storage time. Zeta behavior aligned with the crispness behavior in the tests of increasing and decreasing crispness over time.

Keywords: Crispness, Convolutional Neural Network, Librosa, Toast, Food Materials.

RESUMO

LOPES, R. Z. **Avaliação da crocância sensorial de alimentos crocantes secos por redes neurais convolucionais**. 2023. 67 f. Dissertação (Mestrado) – Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, 2023.

A secagem convectiva é tradicionalmente utilizada para desidratar alimentos, a fim de reduzir o volume e a atividade de água, possibilitando o fácil transporte e armazenamento. Durante a secagem, os alimentos sofrem redução de volume de acordo com a perda de umidade, resultando em alterações na matriz sólida e formação de estrutura crocante quando esmagados ou fraturados. Este trabalho focou-se em desenvolver métodos de quantificação e classificação de alimentos secos crocantes, tais como batatas chips, torradas e alimentos fritos, como batatas fritas e frango frito. Os perfis de compressão e ruído sonoro foram determinados por um dispositivo de alavanca manual coberto por uma caixa de supressão de ruído. O som capturado foi transformado em diferentes parâmetros com o auxílio de bibliotecas em Python e Mathematica Wolfram. O espectro de potência do sinal sonoro foi obtido pelo método de transformada discreta de Fourier em Wolfram, enquanto o Onset Strength e os coeficientes cepstrais de frequência Mel (MFCC) foram obtidos por meio da biblioteca Librosa. Os espectros sonoros, Onset Strength e MFCC foram processados em redes neurais com o objetivo de classificar a crocância do frango frito, das batatas chips e das torradas. Os modelos de classificação que utilizaram como entradas os sinais DFT e MFCC apresentaram acurácia superior a 95%. Este estudo permitiu descrever o som crocante por meio da intensidade e duração do sinal. Um segundo estudo utilizou código Python e a biblioteca Librosa na tentativa de gerar um número adimensional para classificar a intensidade da crocância, denominado valor Zeta. O valor Zeta foi obtido a partir dos valores de Root Mean Squared Energy, multiplicados pelos picos de intensidade em intervalos de 1 segundo. A validação experimental do valor Zeta foi realizada por meio da aquisição de ruídos de crocância para torradas e batatas fritas, monitorando-se a umidade e o tempo de estocagem. O comportamento de Zeta alinhou-se com o comportamento da crocância nos testes de aumento e diminuição da crocância ao longo do tempo.

Palavras-chave: Crocância, Rede Neural Convolucional, Librosa, Torrada, Materiais Alimentícios.

List of Illustrations

List of Tables

TABLE OF CONTENTS

CHAPTER 1

General Introduction on Crispness Classification and Quantification, Objectives, and Master's Thesis Structure

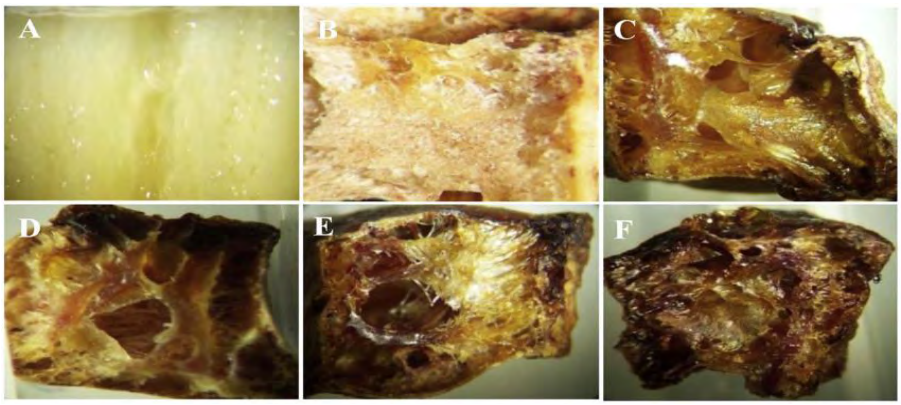
1 INTRODUCTION

1.1 CRISPNESS

The crispness effect, obtained from the formation of a rigid crust on the food, provides highly palatable sensory characteristics (LA FUENTE; LOPES, 2018; MONTEIRO; CARCIOFI; LAURINDO, 2016). The crispy materials obtained by the drying process have similar quality attributes to freeze-dried or fried materials (LA FUENTE; LOPES, 2018). However, drying has advantages such as low processing costs compared to freeze-drying, lower oil content, and longer shelf-life (MONTEIRO; CARCIOFI; LAURINDO, 2016; BI et al., 2015).

There are many methods for achieving the crispy effect in drying trials, such as high-temperature intermittent drying (HTST). The initial high temperature promotes the formation of an initial layer on the surface of the solid; this partially dried layer promotes the "*puffing*" effect (VARNALIS; BRENNAN; MACDOUGALL, 2001). "*Puffing*" involves the release or expansion of vapor or gas within the product, either to create an internal structure or to expand, or rupture an existing one (PAYNE; TARABA; SAPUTRA, 1989). It results in crispy products with good rehydration capacity and faster drying (HOFSETZ et al., 2007; SACA; LOZANO, 2007). Figure 1.1 demonstrates the "puffing" effect from banana slices in a drying cycle.

Figure 1.1. Photographs of banana slices submitted to different drying cycles at high temperatures, it highlights the increasing crispness.



Source: LA FUENTE (2018).

Drying of materials composed of natural polymers can be achieved by various technologies, such as spray drying, freeze drying, fluidized bed, and convective drying. The operation parameters, such as vapor pressure, temperature, air velocity, and equilibrium moisture, need to be monitored to produce a dry solid with specific mechanical properties and

without degradation of active compounds. Dry solids also have easy handling, smaller storage volumes, and reduced transportation cost (MUJUMDAR, 2006).

Convective drying involves the removal of moisture by the simultaneous transfer of heat and mass between the moist solid and the drying air. Energy in the form of heat is transferred to the moist solid by the hot air stream. Vaporization of the liquid occurs at wet bulb temperature at the surface of the moist solid, and heat transfer may be due to convection, conduction, or radiation (CASTRO; MAYORGA; MORENO, 2018). Additionally, during drying, the decrease in moisture of the materials can also result in the modification of mechanical properties and an increase in the modulus of elasticity. As a result, drying modifies the mechanical strength of these materials, which can be evaluated through the variation of three sound parameters: energy, intensity, and continuity.

Crispness obtained by the fragmentation of dry food is a property little studied in food engineering (SPENCE, 2016). Machine learning and Deep Learning demonstrate a new horizon for food analysis, with few published works in the world and a high potential for impact in the area if well-directed. Thus, this research project established methods that made it possible to describe changes in the sound properties of food materials. We also evaluated how this behavior varies over time and frequency by comparing spectrograms and intensity plots over drying times.

1.2 LIBROSA AND NEURAL NETWORKS APPLICATIONS ON CRISPNESS EVALUATION

Artificial Neural Networks are a branch of Deep Learning, whose first mathematical model was conceived from observations and hypotheses about the biological behavior of the neural system (MCCULLOCH; PITTS, 1943). In the year 1948, the researcher Donald Hebb succeeded in finding a method of neuron training based on the neurophysiology of nerve cells (MORRIS, 1999). The early models had limitations in classifying databases into more than two classes, and the relationship between the data had to be linear, which was impractical for the more complex applications (DA SILVA et al., 2017).

One of the most complete and universal architectures for application in classification and pattern discovery is the Convolutional Neural Network (CNN). It separates, for example, an image into several parts, and a neuron is assigned to analyze the information and send it to the adjacent layer of neurons that connects all the information from each convolution to reach a common output (GU et al., 2018; ZHOU, 2020). For a Neural Network to be trained, the data needs to go through a process of acquisition, pre-processing, and feature extraction to direct the

learning to what the network designer wants. This is why raw crispness audio is not suitable for a CNN. Mel-Frequency Cepstral Coefficients (MFCC) have shown great potential in classification by transforming sound into fundamental frequencies known as Formants (JI et al., 2021). Formants represent the information needed for a human being to be able to distinguish sounds, even if they are as similar as the crispness of a Fried Chicken and a piece of toast.

One practical approach to sound processing is through Librosa, a Python library specialized in sound signal and music processing that provides a bunch of features like MFCC and Onset Strenght. The former is defined through the frequency domain while the latter is a relationship between wave amplitude and time, i.e., it is a versatile and powerful tool (MCFEE et al., 2015; RAGURAMAN; R.; VIJAYAN, 2019). The Librosa library can be used to extract MFCCs from sound excerpts to input them into CNN for music genre classification (CHENG et al., 2021).

One of the first studies to analyze neural networks in food differentiated the acoustic characteristics of 5 types of crispy snacks reaching 89% accuracy in the Neural Network. (LIU; TAN, 1999). Crispness is correlated with intonation; the sound is sharp and short similar to a fabric ripping. Its quality perception is interrelated with energy and loudness, the sharper and more energetic the sound, the greater the consumer acceptance (TUNICK et al., 2013; VICKERS, 1984). Other studies in the area use neural networks to estimate the texture of various crispy foods using fracture tests and acoustic tests on texturometers. (CHEN; DING, 2021; KATO et al., 2018, 2019). The data is collected by placing a microphone near the region where the food is sheared.

The Librosa library is an audio analysis and processing package capable of isolating key parameters such as Onset Strength, MFCC, and Beat that measure the intensity, frequency, and amplitude of the sound noise, which is essential to identify crunchiness patterns in the foods that will be studied (LIBROSA, 2021). A person defines a food as crunchy at the moment of its first bite, so we will also study if the noise coming from the bite of a crunchy food is present in the percussive or harmonic spectrum and what is the intensity of this sound. The graphics generated in Librosa are sound spectrograms that transform sound into spectrograms and waveplots, which facilitate the identification of patterns.

Keras is a machine-learning library that will be used to validate the patterns identified in Librosa by simulating a human brain that trains by learning what is and is not crunchy (CHOLLET, 2017). The number of neurons and layers depends on the complexity of the problem, in the simplest case, there is one input, the audio data, and two outputs, crispy and not

crispy. In a more complex case, there will be four outputs each with a crispness profile such as hard or very crispy, ideal crispness, not very crispy, and not crispy to be compared with a crispness sensory analysis test.

1.3 OBJECTIVES

This work focused on evaluating crispness by its sound only and developing a dimensionless number, Zeta, that describes its behavior over time. Crispness is a mixture of mechanical and sound characteristics, but customer decision is mostly based on the sound. It is a forgotten flavor, which intensifies a product's quality. Furthermore, this work proposed the best Zeta range for French Fry and Bread, which indicates the maximum peak for crispness. Finally, this study created a marketable software that calculates Zeta from an audio input. It can be updated with new Zeta ranges and be an easy-to-use method for laboratories.

The specific objectives of this Master's Thesis were:

- Organize two small sound databases. One with random ASMR crispy audios from YouTube, it featured audios of Toast, Fried Chicken, and Potato Chips. The second centered on toast audios from a 300-minute drying process.
- Identify which sound parameters best represent crispness.
- Successfully classify the ASMR audios while evaluating which Neural Architecture better fit the chosen parameters.
- Perform drying experiments to analyze how the sound parameters behave over time.
- Achieve a dimensionless number, Zeta, which best represents crispness behavior.
- Test Zeta in a Control Quality Laboratory from a big company, Ingredion.
- Refine Zeta after evaluating its results.
- Create an easy-to-use software to calculate Zeta.

1.4 MASTER'S THESIS STRUCTURE

The dissertation was elaborated in three chapters as follows:

Chapter 1 presents a general introduction to crispness, analysis of temporal and spectral parameters of the crispy sound, and the applied neural networks. The research objectives and justification conclude the chapter.

Chapter 2 presents the studies in the classification of crispy sounds coming from random internet videos. The goal of this chapter is to find the parameters that best describe the differences in crispness.

Chapter 3 presents the studies in crispness quantification. The chapter focused on developing a dimensionless number for crispness, which was done by correlating the essence of the parameters found in Chapter 3.

Chapter 4 presents the final remarks.

REFERENCES

BI, J. et al. Evaluation indicators of explosion puffing Fuji apple chips quality from different Chinese origins. **LWT - Food Science and Technology**, v. 60, n. 2, p. 1129–1135, mar. 2015.

CASTRO, A. M.; MAYORGA, E. Y.; MORENO, F. L. Mathematical modeling of convective drying of fruits: A review. **Journal of Food Engineering**, v. 223, p. 152–167, 2018.

CHEN, L.; DING, J. Analysis on Food Crispness Based on Time and Frequency Domain Features of Acoustic Signal. **Traitement du Signal**, v. 38, n. 1, 2021.

CHENG, Y.-H. et al. Automatic Music Genre Classification Based on CRNN. **Engineering Letters**, v. 21, n. 1, p. 312–316, 2021.

CHOLLET, F. Deep Learning with Python. Manning Publications, 2017.

DA SILVA, I. N. et al. **Artificial Neural Networks**. Cham: Springer International Publishing, 2017.

GU, J. et al. Recent advances in convolutional neural networks. **Pattern Recognition**, v. 77, p. 354–377, 2018.

HOFSETZ, K. et al. Changes in the physical properties of bananas on applying HTST pulse during air-drying. **Journal of Food Engineering**, v. 83, n. 4, p. 531–540, 2007.

KATO, S. et al. **Snack Food Texture Estimation by Neural Network**. 2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS). **Anais**...IEEE, 2018

KATO, S. et al. Snack Texture Estimation System Using a Simple Equipment and Neural Network Model. **Future Internet**, v. 11, n. 3, 2019.

LA FUENTE, C. I. A.; LOPES, C. C. HTST puffing to produce crispy banana - The effect of the step-down treatment before air-drying. **LWT**, v. 92, n. November 2017, p. 324–329, jun. 2018.

LIBROSA. Biblioteca do Librosa. 2021 Disponível em:

<https://librosa.org/doc/latest/index.html> Último acesso em: <14/01/2021>

LIU, X.; TAN, J. ACOUSTIC WAVE ANALYSIS FOR FOOD CRISPNESS EVALUATION. **Journal of Texture Studies**, v. 30, n. 4, 1999.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The Bulletin of Mathematical Biophysics**, v. 5, n. 4, p. 115–133, dez. 1943.

MCFEE, B. et al. **librosa: Audio and Music Signal Analysis in Python**. 2015.

MONTEIRO, R. L.; CARCIOFI, B. A. M.; LAURINDO, J. B. A microwave multi-flash drying process for producing crispy bananas. **Journal of Food Engineering**, v. 178, p. 1–11, jun. 2016.

MORRIS, R. G. . D.O. Hebb: The Organization of Behavior, Wiley: New York; 1949. **Brain Research Bulletin**, v. 50, n. 5–6, p. 437, nov. 1999.

MUJUMDAR, A. **Handbook of Industrial Drying, Third Edition**. [s.l.] CRC Press, 2006.

PAYNE, F. A.; TARABA, J. L.; SAPUTRA, D. A review of puffing processes for expansion of biological products. **Journal of Food Engineering**, v. 10, n. 3, p. 183–197, jan. 1989.

RAGURAMAN, P.; R., M.; VIJAYAN, M. **LibROSA Based Assessment Tool for Music Information Retrieval Systems**. 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). **Anais**...IEEE, 2019

SACA, S. A.; LOZANO, J. E. Explosion puffing of bananas. **International Journal of Food Science & Technology**, v. 27, n. 4, p. 419–426, jul. 2007.

SPENCE, C. **Sound: The Forgotten Flavor Sense**. [s.l.] Elsevier Ltd, 2016.

TUNICK, M. H. et al. Critical Evaluation of Crispy and Crunchy Textures: A Review. **International Journal of Food Properties**, v. 16, n. 5, 2013.

VARNALIS, A. I.; BRENNAN, J. G.; MACDOUGALL, D. B. A proposed mechanism of high-temperature puffing of potato. Part I. The influence of blanching and drying conditions on the volume of puffed cubes. **Journal of Food Engineering**, v. 48, n. 4, p. 361–367, jun. 2001.

VICKERS, Z. M. CRISPNESS AND CRUNCHINESS - A DIFFERENCE IN PITCH? **Journal of Texture Studies**, v. 15, n. 2, 1984.

ZHOU, D.-X. Universality of deep convolutional neural networks. **Applied and Computational Harmonic Analysis**, v. 48, n. 2, p. 787–794, mar. 2020.

CHAPTER 2

Food Crispness Classification by Deep Neural Networks

2 FOOD CRISPNESS CLASSIFICATION BY DEEP NEURAL NETWORKS

Rafael Z. Lopes, Gustavo C. Dacanal*

Department of Food Engineering, Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, FZEA-USP, 13635-900, Pirassununga, SP, Brazil

* Corresponding author: Gustavo C. Dacanal, E-mail gdacanal@usp.br, Tel/Fax +55 (19) 35654284, ORCID 0000-0002-6061-0981

ABSTRACT

Crispness is a textural characteristic that influences consumer choices, requiring a comprehensive understanding for product customization. Previous studies employing Neural Networks focused on acquiring audio through mechanical crushing of crispy samples. This research investigates the representation of crispy sound in time intervals and frequency domains, identifying key parameters to distinguish different foods. Two machine learning architectures, Multi Layer Perceptron (MLP) and residual neural network (ResNet), were used to analyze Mel Frequency Cepstral Coefficients (MFCC) and Discrete Fourier Transform (DFT) data, respectively. The models achieved over 95% accuracy "in-sample" successfully classifying fried chicken, potato chips, and toast using randomly extracted audio from ASMR videos. The MLP (MFCC) model demonstrated superior robustness compared to ResNet and predicted external inputs, such as freshly toasted bread acquired by a microphone or ASMR audio of toast in milk. In contrast, the ResNet model proved to be more responsive to variations in DFT spectrum and unable of predicting the similarity of external audio sources, making it useful for classifying pre-trained “in-samples”. These findings are useful for classifying crispness among individual food sources. Additionally, the study explores the promising utilization of ASMR audio from Internet platforms to pre-train Artificial Neural Network (ANN) models, expanding the dataset for investigating the texture of crispy foods.

KEYWORDS: Convolutional Neural Networks, Mathematical Modeling, Crispness, Fried Chicken, Potato Chips, Toast.

2.1 INTRODUCTION

The use of Artificial Neural Networks has increased during the last decades demonstrating feasible results in several areas from engineering to health, an example is the

analysis of babies' cries to identify serious diseases (Ji, Mudiyanselage, Gao, & Pan, 2021). Crispness is one of the most relevant characteristics when buying a product, but it is only evaluated through mechanical properties such as texture. This is an opportunity to impact the food analysis market by uncovering the characteristics of the crispy sound (Buisson & Silberzahn, 2010).

Crispness is a sensory attribute related to food texture and its sound perception is extremely important for the purchase decision and quality perception by consumers, since it indicates product freshness (Lawless & Heymann, 2010). The sound events of crispy dry food occur due to the structure breaking sound and the release of air. When applying force with the incisor teeth, energy is retained and dissipated in the form of sound energy during rupture (Dias-Faceto, Salvador, & Conti-Silva, 2020).

Studies have been conducted to correlate sound crispness and mechanical crispness by evaluating the sound in time domain, acoustic signal amplitude, duration and number of peaks (Akimoto, Sakurai, & Blahovec, 2018; Dias-Faceto et al., 2020; Gouyo et al., 2020; O'Shea & Gallagher, 2019), but, nowadays, with the use of artificial intelligence it is possible to predict the crispness of food more quickly, accurately and advantageously by performing sound analysis (Chen & Ding, 2021; Liu, Cai, et al., 2021b).

Foods that produce sounds when sheared by biting are known as crispy foods, a niche product within the Food Engineering spectrum. The sound characteristics mostly come from the frying, baking, and roasting process. When their water activity has been decreased, air-filled voids appear in their structure. They are responsible for the better propagation of the sound when eating. Studies indicate that the difference between the sounds of each crispy food is in the intonation, where crispness is categorized as a higher-pitched sound while crunchiness is a lower-pitched sound (VICKERS, 1984). If it is possible to identify differences in sound empirically, it is also possible for a neural network to classify the crispness.

Every crispy food has a sound, they may vary depending on the food, chips have smaller ranges than toast. Training an artificial neural network to classify crispy foods is the first step. Studies in this area approximated the texture function using Force data from texturometers (Kato, Ito, Wada, Kagawa, & Yamamoto, 2018; Kato et al., 2019a; Tunick et al., 2013). They had challenges regarding the equipment's noise, a small change in the sound caused deviations in the results (Andreani et al., 2020; de Moraes et al., 2022). A proposed solution developed a swing arm device capable of capturing a cleaner sound, they approximated the energy using the friction force (Akimoto et al., 2018). The approach taken in this work utilized ASMR audios processed in python, there are no forces involved, just the sound itself to be evaluated.

Crispness Classification

The Librosa library is an audio analysis package capable of isolating key parameters such as Mel Frequency Cepstral Coefficients (MFCCs), which unfold the sound's identity. Keras is a machine learning library that will be used to validate the patterns identified in Librosa by simulating a human brain that trains by learning which crispness is from which food (Chollet, 2017). A Kaggle challenge inspired the first machine learning architecture. They developed a simple fully connected deep neural network to classify eating sounds of 20 different types of food. (Ma, Gómez Maureira, & van Rijn, 2020) The best model won using the MFCC as the input, which brought a 90% accuracy.

The Convolutional Neural Network is a more complex architecture, it's capable of handling large amounts of data. Their application is almost universal but more focused on image and audio classification (Zhou, 2020). The convolution process is a multiplication operation between the terms of two arrays, the original and a kernel filter, resulting in a smaller matrix. The two developed models received the Discrete Fourier Transform (DFT) linear data and spectrogram images as their input, respectively.

The application of artificial neural networks (ANNs) in assessing food texture, particularly crispness, has been explored in various studies, as shown in Table 2.1. These networks, including Back Propagation Neural Networks (BPNN), Feedforward Neural Networks (FNN), and Multi-Layer Perceptrons (MLP), have been used to analyze acoustic signals generated during mechanical tests on food samples. The frequency range of these signals varies, but often falls within 0-20 kHz (Chen & Ding, 2021; Iliassafov & Shimoni, 2007; Kato et al., 2018, 2019a, 2019b; LIU & TAN, 1999; Liu, Cai, et al., 2021a; Liu, Wu, et al., 2021; Przybył, Duda, Koszela, & Stangierski, 2020; Sanahuja, Fédou, & Briesen, 2018; Srisawas & Jindal, 2003; Wietlicka, Muszyński, & Marzec, 2015).

For instance, Chen et al. (2021) used a BPNN model to analyze the crispness of vegetables like potatoes and carrots, while Iliassafov et al. (2007) used a similar model to predict the sensory crispness of coated turkey breasts. Kato et al. (2018, 2019a, 2019b) employed a simple BPNN model to quantify the texture of snacks such as rice crackers and potato chips. Liu et al. (1999) used a FNN to evaluate the crispness of Chex Mix products, and Przybył et al. (2020) utilized MLP to analyze the quality of dried strawberries.

In addition to crispness, these models have been used to assess other food qualities. For example, Liu et al. (2021a, 2021b) used a BPNN model for non-destructive evaluation of apple firmness. Sanahuja et al. (2018) and Srisawas et al. (2003) used ANNs to classify the freshness of puffed snacks and the moisture content of snack foods, respectively. Wietlicka et al. (2015) used ANNs to classify extruded bread samples based on acoustic emission signals.

Crispness Classification

Table 2.1: Cited publications, correlating "(crispy or crunch or crispness or crunchiness) and (neural network) and (sound or acoustic or audio or signal) and (food)" research from 2003 to 2023.

Author (year)	ANN architecture	Samples	Audio source	Type of Signal (input: output neurons)	Power spectrum frequency range
CHEN et al. (2021)	BPNN	Potato, sweet potato, carrot, and turnip	Experimental (Compression in mechanical tests)	Waveform index; peak PSD amplitude; and amplitude difference (3:1)	(0-20 kHz) not detailed
ILIASSAFOV et al. (2007)	BPNN	Coated turkey breast (frying, oven, and microwave)	Experimental (Compression by a texture analyzer)	FFT (7:3)	(0-8 kHz)
KATO et al. (2018; 2019a, 2019b)	Simple neural network model (BPNN)	Rice crackers; Potato chips; Wafers; Cookies; Biscuits; Corn snacks	Experimental (Stacked and crushed by the equipment, pair of pincers or pliers)	FFT with five integration parts (10:2)	(0-2 kHz) or (0-4 kHz) or (0-10 kHz)
LIU et al. (1999)	Feedforward neural network (FNN)	Corn chex; Wheat chex; Round pretzel; Rye chip; and Bread twist	Experimental (Crushed by a pair of pliers)	STFT or power spectrum (5:3)	(0-20 kHz)
LIU et al. (2021a, 2021b)	BPNN	Apple small cuboid	Experimental (Crushed by a steel ball knocking)	FT or HHT (24 neurons; not detailed)	(0-10 kHz)
PRZYBYŁ et al. (2020)	MLP	Dried strawberries	Experimental (Falling into water, or crushed by a texture analyzer)	Frequency and sound intensity (2:1)	(0-16 kHz)
SANAHUJA et al. (2018)	BFFN	Puffed snacks under controlled RH	Experimental (Crushed by a texture analyzer)	STFT or CWT or HHT (Input with 68 features as 1/3 octave bands; not detailed)	(0-20 kHz)
SRISAWAS et al. (2003)	BPNN or PNN	Pringles brand potato chips; Paprika brand extruded snacks; Munchy brand crackers	Experimental (Cutting with a pair of pincers, imitate biting with incisors)	FFT (102:1 or 102:4)	(0-7 kHz)
ŚWIETLICKA et.al. (2015)	RBF or SOM	Extruded flat graham; corn; and rye breads at different water activity levels	Experimental (Compression plate)	Acoustic emission in relation of bread type, water activity value, or both (4:3)	(0-22 kHz) not detailed
LOPES & DACANAL (2023) *"This work"*	ResNet or MLP	Fried chicken, Potato chips, Toast and Toast in milk	Internet platform (ASMR mastication videos); and Experimental validation	DFT or MFCC coefficients (100:3 or 64:3)	(0-11 kHz)

These studies demonstrate the potential of ANNs in food quality assessment, offering valuable insights for the food industry. However, the complexity of these models and the need for further optimization highlight the ongoing challenges in this field, which is the contribution of this work.

As presented in Table 2.1, the predominant type of signal employed in artificial neural network (ANN) architectures is derived from Fourier transforms, namely FFT (Fast Fourier Transform), STFT (Short-Time Fourier Transform), or DFT (Discrete Fourier Transform). Furthermore, other signal types such as Hilbert-Huang Transform (HHT), Continuous Wavelet Transform (CWT), and waveform signals are also utilized. In this study, we utilize the signal MFCC (Mel Frequency Cepstral Coefficients), which has not been explored in the literature for crispness analyses.

This paper aims to investigate the unique characteristics of the crispy sound and its differences for each of the three presented foods: Fried Chicken, Potato Chips and Toast. The sound influences how many people perceive the desirable characteristics of a food, yet they forgot how it can change the flavors and sensations in their consumption (Spence, 2016). Understanding their characteristics and differences allows one to optimize their manufacturing processes in search of the most appealing sound to the consumer. However, it should first be explored whether it is possible to generalize a function that describes the crispness behavior by comparing the sounds of different foods in the two proposed neural network models.

2.2 MATERIAL AND METHODS

2.2.1 Overall scheme for audio acquisition, preprocessing, and spectrum generation

A total of 584 digital audio files collected from a web platform (YouTube, 2022) were the main source of three different classes: Fried Chicken, Potato Chips, and Toast. The selected "*ASMR*" category provided clean videos.

Audacity® audio editing software was used to trim the audio files to a duration of 1 second. The resulting audio files were then converted to monophonic audio and their sampling rate (SR) was standardized to 22050 Hz. The 584 audio files were exported in the .wav file format.

The use of these 1s-length segments followed the "Fair use on *Youtube*" Guidelines (http://www.support.google.com/youtube/answer/9783148). Furthermore, there was no exposure of the recording owners and any unauthorized use of it.

Wolfram Mathematica and Python were the two main languages used for processing and extracting spectrum data from the audio files, as shown in Figure 2.1. Two types of deep

neural networks were used as a case study for the classification of the crispness of Fried chicken, Potato chips, and Toast: Residual Neural Network, and Multilayer Perceptron Network. The details are provided below.

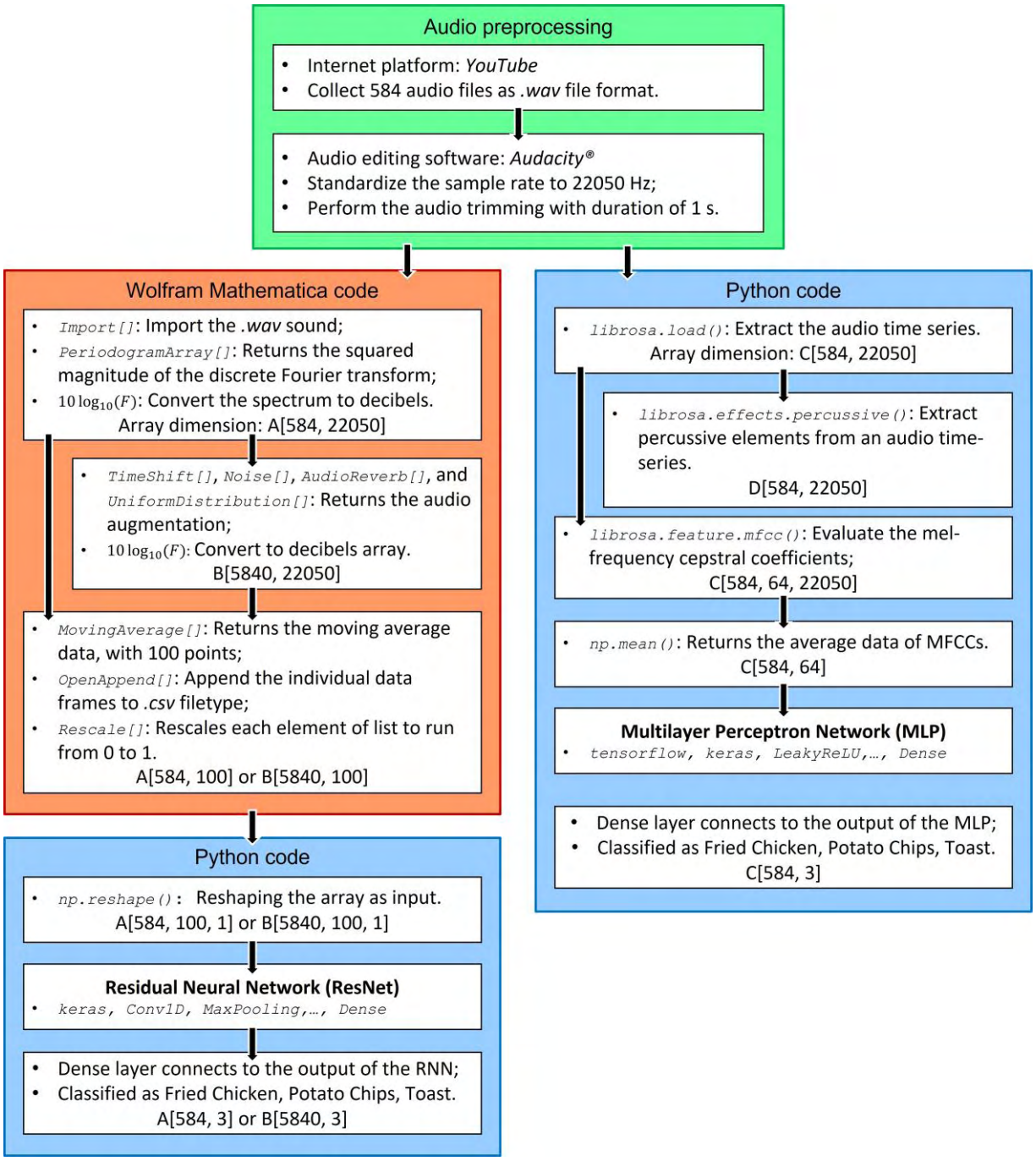
The inputs for the Residual Neural Network (ResNet) were obtained using Wolfram, which evaluated the Discrete Fourier Transform (DFT) as a preprocessing step. An additional step was taken to rescale the spectrum from 0 to 1. This resulted in DFT spectrum arrays with 584x100 terms, which were imported to Python using the "*.csv*" file format and *pandas* library. In a second study using the ResNet network, audio augmentation filters were applied to enlarge the array of DFT spectrum inputs to 5840x100 terms. To input the ResNet, the structured data required an additional dimension and reshaping the array. This dimension was then transformed into the dimension of the filter.

The inputs for the Multilayer Perceptron Network (MLP) were obtained using Python code with the *librosa* library. Specifically, *librosa.load* generated an audio time series, while *librosa.feature.mfcc* was used to evaluate the Mel-frequency cepstral coefficients (MFCCs). The resulting MFCC arrays contained 584x64 terms. In the Librosa library v0.9.0, the default number of MFCCs is 20. However, in this study, the neural network achieved its best performance using 64 MFCCs.

Furthermore, the *librosa.effects.percussive* function decomposed the audio signal into percussive components, which were necessary for a more comprehensive analysis of the crisp sound.

Through pre-testing and adjusting the input array sizes in the ResNet and MLP neural networks, the optimal dimensions were determined to be 100 and 64, respectively. These values were selected based on the resulting improvements in model predictions observed during testing. The model accuracy is typically used as the primary metric for evaluating the performance of the model during training and testing. The goal of evaluating model accuracy is to ensure that the model can make accurate predictions on new data. The model loss measures the difference between the predicted outputs and the true labels for a set of training examples, and it is the quantity that the model is trying to minimize during training. Both model accuracy and model loss were utilized to monitor the training and predictive capabilities of the model.

Figure 2.1. Audio processing via Python and Wolfram Mathematica.



2.2.2 DFT spectrum for ResNet model

The Mathematica Wolfram code provided performs some operations on original and augmented crispy audios and produces a power spectrum plot of the audio. The audio file is first mixed into a mono channel and normalized. The AudioPlot function is used to plot the audio signal as a function of time. Then, the periodogram is computed with 200 points, and the length of the periodogram is used to calculate the frequency range of the spectrogram. The frequency range is halved, and the periodogram is converted to decibels. The resulting power

spectrum is rescaled to values between 0 and 1, with array size 584x100 or 5840x100, and used as input in the ResNet.

2.2.3 Audio data augmentation for ResNet model

As previously mentioned, audio augmentation was employed as a supplementary study of the ResNet. This step increased the size of the original dataset by a factor of 10, resulting in an array dimension of 5840x100. Wolfram Mathematica utilized four augmentation methods to generate additional audio data for the ResNet: *Timeshift*, *Reverb*, *UniformDistribution*, and *Noise*.

Timeshift was used to shift a portion of the audio signal along the time axis, allowing for the collection of small audio fragments that could be used to generate the DFT spectrum and input into the ResNet. *Reverb* was used to apply reverberation to the sound signal, producing echoes and slightly altering the original data. *UniformDistribution* was used to randomly vary the amplitude of the audio signals uniformly. *Noise* was applied to all datasets as a way of introducing random variations to the audio data.

2.2.4 Mel-frequency spectrogram and MFCC coefficients for MLP model

MFCCs are based on Mel Frequency and Cepstrum. Mel Frequency estimates pitch logarithmically, and Cepstrum is a spectrum of a spectrum. Equation 2.1 transforms a time-domain signal into a log amplitude spectrum using Fourier Transform and the inverse Fourier transform, separating the information relative to the Formants. MFCCs use Mel-Scaling and Discrete Cosine Transform instead of the inverse Fourier transform.

$$C\big(x(t)\big) = F^{-1}[\log(F[x(t)])] \tag{2.1}$$

Where C is the Cesptrum, $x(t)$ is the time-domain function, and F is the Fourier transform.

The Python code used *librosa.feature.mfcc* to perform the MFCC calculation after loading the audio, resulting in an MFCC array. The number of MFCCs was set to 64 instead of the default 20 and scaled to fit the neural architecture. The transposed MFCC array and *np.mean* function provided 64 scaled MFCC values for the MLP architecture without averaging.

The Mel spectrogram, created using the Librosa library, applies the Mel Scale to show the correlation between energy, frequency, and time, similar to the MFCC calculation without the Discrete Cosine Transform. The *librosa.display.specshow* function applies the Mel Scale to the y-axis, enabling comparison of sound crispness in this study.

2.2.5 Beat detection in crispy sounds

The Beat feature is a useful tool for visualizing the changes in crisp sound amplitude over time in the time domain. Its use in conjunction with a filtered Mel Spectrogram can reveal the rhythmic structure of the sound, including the onset and offset times of individual bites. The *librosa.beat.beat_track* function in Python employs dynamic programming (ELLIS, 2007) to calculate the beat. The code first reads audio files from a directory and calculates their duration using the Librosa package. It then generates a summary of the total number of audio files per class, a box plot displaying the distribution of audio file durations, and the mean and variance of audio file durations. By correlating sound amplitude with time, this method can estimate the locations of the beat. For musicians, the beat represents the basic unit of measurement for melody and reflects the speed at which the music is played (LEVITIN, 2007).

2.2.6 Residual network architecture (ResNet)

The residual learning process involves stacking multiple layers into a block and then adding the result of the block to its first layer. This approach prevents a decrease in training accuracy when constructing complex structures. This procedure is widely used in deep learning and is effective in improving model performance (HE et al., 2015).

The ResNet study was a Python script that uses TensorFlow and Keras libraries to create a convolutional neural network (CNN) model for audio classification. The model is based on the ResNet architecture, which uses residual blocks to improve training performance.

The script reads the spectrum data that was previously evaluated by Wolfram. The ResNet's data input contained the DFT audio spectrum, with an array dimension of 584x100, or 5840x100 if audio augmentation was used. The output data used in ResNet contained binary classification for crispy sources: Fried Chicken, Potato Chips, and Toast, with an array dimension of 584x3, or 5840x3 if audio augmentation was used. For example, an array line indicating Fried Chicken was represented as (1, 0, 0), while Potato Chips as (0, 1, 0), and Toast as (0, 0, 1). The audio data was split into training and validation sets and then used as input and output for ResNet.

Figure 2.2 shows the main function that defines the architecture of the ResNet-based CNN model, which consists of a series of residual blocks, followed by average pooling and dense layers.

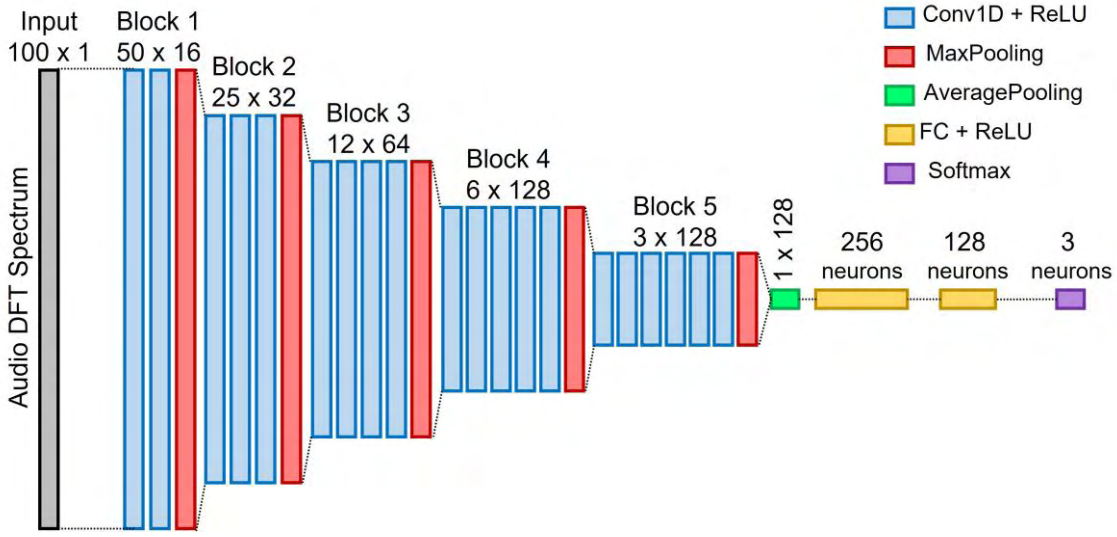
The compile method is utilized to configure the model for training, specifying the Adam optimizer and categorical cross-entropy loss function. The chosen hyperparameters were

ADAM optimizers with a learning rate of 0.001 and a categorical cross-entropy loss function. A batch size of 50 was selected, and dropout was not applied.

Callbacks were used to monitor the training progress and save the best model. Specifically, the EarlyStopping callback is used to stop training when the model is not improving, and the ModelCheckPoint callback is used to save the model with the highest validation accuracy.

Finally, the fit method is used to train the model on the training data, using a batch size of 32 and a total of 100 epochs. The model is evaluated on the validation set after each epoch, and the training progress was displayed.

Figure 2.2. ResNet architecture for an audio DFT spectrum input.



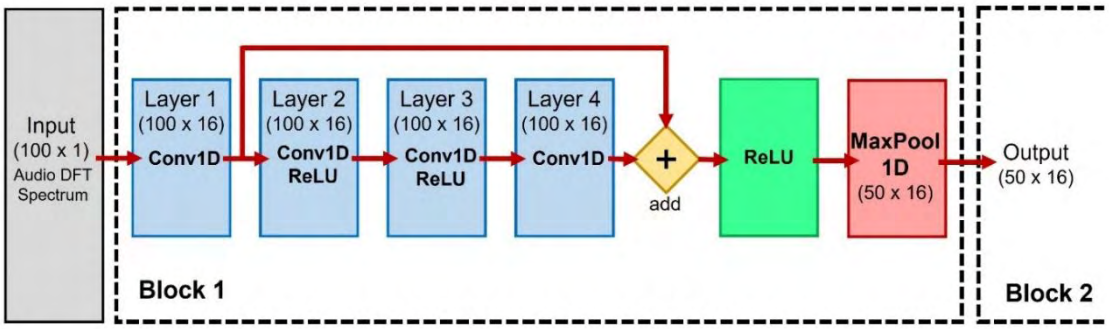
The residual blocks 1, 2, 3, 4, and 5 were constructed using a sequence of Conv1D, ReLU, and MaxPooling layers. This procedure enabled the creation of layers with reduced dimensions, allowing them to be connected to the dense layers and the output array containing the binary classification of crispy sounds.

As an example, Figure 2.3 illustrates the detailed construction of Block 1. The audio spectrum input, with 100x1 terms, is passed through a sequence of four layers of Conv1D, along with the ReLU activation function and MaxPooling. As a result, the residual block changes the original array dimension to 50x16 terms. After filtering over the sequence of residual blocks, the original data is progressively reduced until it is connected to dense layers composed of 256, 128, and 3 neurons.

The ResNet model was inspired by a study on the classification of damaged structures using impact sound (DORAFSHAN; AZARI, 2020). The authors of that study used the Conv1D

architecture with an input of an image spectrogram. In the present study, we propose a modification by using residual blocks that begin training with an audio DFT spectrum instead of an image. This approach has not been previously explored and may offer improved results in classification tasks involving audio signals.

Figure 2.3. Detailed architecture of residual Block 1 in the ResNet model.



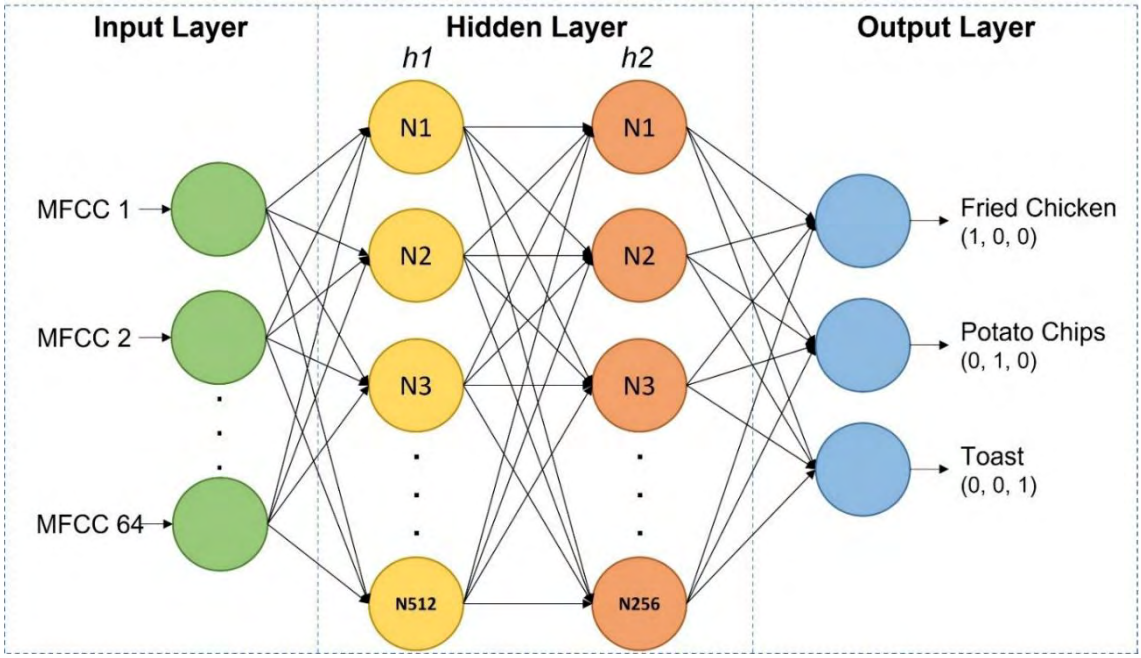
2.2.7 Multilayer perceptron architecture (MLP)

A Multilayer Perceptron (MLP) was employed as the second model for the audio classification of crispy foods, using Mel-frequency cepstral coefficients (MFCCs) extracted from audio signals as input. The MLP is typically composed of an input layer, one or more hidden layers, and an output layer.

Figure 2.4 shows the network architecture comprised of three fully connected (FC) layers, with 512 neurons in the first layer, 256 neurons in the second layer, and 3 neurons in the last layer. Each layer was followed by batch normalization, *LeakyReLU* activation, and dropout layers with a dropout rate of 20%. The model was trained using the categorical cross-entropy loss function and Adam optimizer with a learning rate of 0.001. The MLP input consists of an array containing 64 coefficients of MFCCs that were evaluated for each crisp sound.

The MFCC features were extracted from 584 audio files using the Librosa package and saved as scaled NumPy arrays with their corresponding labels. The data was divided into a training set consisting of 467 audio files and a validation set of 117 audio files, to ensure an appropriate training-validation ratio. The FC audio classification model was defined using the Keras package, compiled, and trained using the training data. The model was trained for 50 epochs, and the model with the lowest validation loss was saved. The training and validation loss and accuracy were recorded over the epochs.

Figure 2.4. MPL architecture for an audio MFCC coefficients input.



2.2.8 Alternatives neural networks

As a brief report, this study attempted to test alternative models for classifying crispness, but none of them provided satisfactory results: LeNet-5 model, and EfficientNetB0 model. Therefore, the ResNet model with augmented data and MLP model with MFCCs were determined to be the most effective approach for accurately classifying crispness based on sound.

LeNet-5 architecture involved only 2 convolutional neuron and its use to classify up to 10 patterns. (Lecun, Bottou, Bengio, & Haffner, 1998) The main change from the original model is the use of Conv1D instead of the two-dimensional convolution. Some adjustments to the number of neurons, kernel size, pool size, and filters were applied to better fit to parameters. This model trained with 3 different input dimensions: the onset strength with 50 parameters, the DFT with 100 parameters, and the audio itself at a sample rate of 2048 Hz. The input parameters came from three different functions: librosa.onset.onset_strength, Mathematica DFT function, and librosa.load at a sample rate of 2048 Hz. The input dimension for each case is 50x1, 100x1, and 2048x1 as they followed the same reshape method as the input for the ResNet. LeNet-5 model has only three Conv1D layers with ReLU, two max pooling layers, and then the fully connected network with a 20% Dropout. The output layer stays the same as before using the softmax function.

EfficientNetB0 is one of the premade models installed in the Keras Library. When it's called in the Collab, it comes as a ready to use model with pretrained weights of Imagenet. This model was created to be an upgrade from the GoogleNet architecture in the Imagenet challenge (Tan & Le, 2019). This model uses the two-dimensional convolution and performs better at classifying images. The input shape has to be fixed at image width, length, and color channels of 224, 224, and 3 respectfully. Therefore, the preprocessing steps are different compared to the Conv1D. Instead of building an array made of the DFT, we had to generate a DFT spectrogram image for each sound to fit the model. The 584 images were divided into 3 different folders in the google drive that were called by the tf.keras.preprocessing.image_dataset_from_directory. This function transformed the 584 images into a dataset that needs to be treated. All the values from this dataset ranges between 0 and 255, therefore all the value needed to be divided by 255 to ensure uniformity of the input..

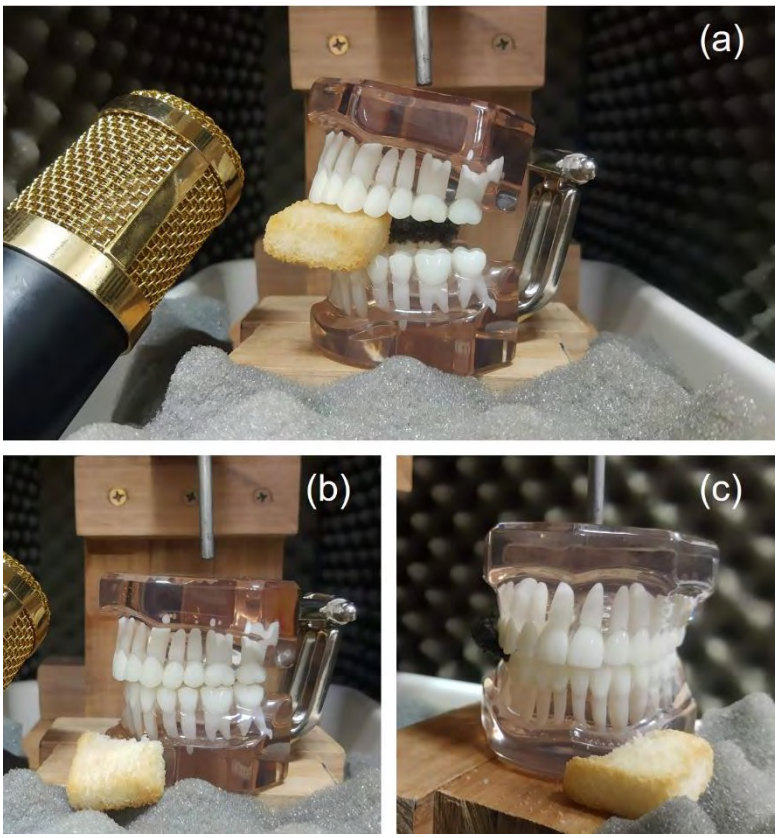
2.2.9 ASMR mastication audios of toast in milk

As part of the cross-validation step for pre-trained ANNs models, ASMR videos were used, featuring individuals eating toast that had been soaked in milk. Audio extraction and the generation of DFT or MFCC signals were carried out, replicating the inputs used during neural network training in section 2.1. The choice of toast immersed in milk aimed to highlight similarities between fresh toast and toast soaked in milk, while assessing the performance of the MLP (MFCC) and ResNet (DFT) models in classifying toast crispness.

2.2.10 Audio acquisition from mechanical crushing of fresh toast samples

In the experimental validation, 50 audio samples of fresh toast (dry) acquired during compression trials were used. These samples were purchased from a local supermarket and bitten by a dental prosthesis. Specifically, the samples were positioned under the pair of premolar teeth and subjected to mechanical load within a soundproof box, as illustrated in Figure 2.5. The audio was captured using a BM800 microphone model at a sample rate of 44.1 kHz, utilizing Audacity software. The previously generated ANN models were used to predict whether the captured audios were of toast or not. This comparison enabled the selection of a preferable model for use with external data source.

Figure 2.5. Audio acquisition through the crushing of toast by a dental prosthesis in a soundproof box.



2.2.11 Cross-validation of pre-trained ANNs

Cross-validation of pre-trained ResNet and MLP models was used to test the ANN's performance on external inputs and verify the similarity between the predicted results and the expected results. This validation set typically consists of data that was not used during the training process. By providing these external inputs to the pre-trained ANN and comparing the predicted results with the known expected results, the model assessed the similarity of the predictions.

The percentage values representing the crispness predictions made by pre-trained Artificial Neural Networks (ANNs) for two different conditions: fresh toast and toast in milk. These predictions indicate whether the crispness values estimated by the ANNs (Fried Chicken, Potato Chips, and Toast) were similar or not.

2.3 RESULTS AND DISCUSSION

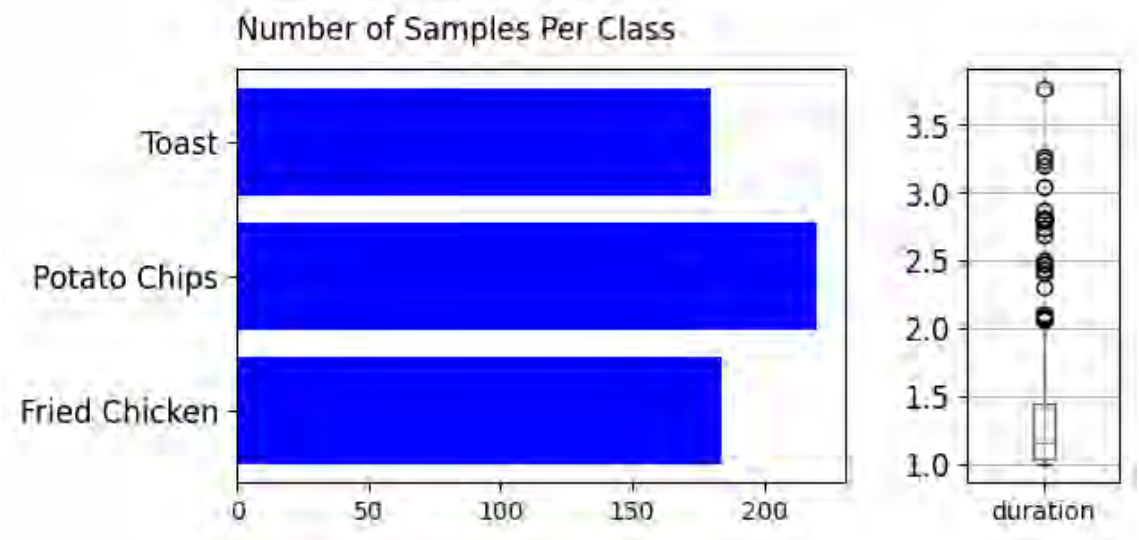
2.3.1 Preprocessing analysis

Deep learning projects often begin with a hypothesis, a challenge, or a curiosity. For instance, one may wonder whether every crisp sound is distinct and what implications this knowledge may hold for future projects. This study aims to address these questions.

The decision to use three distinct types of crispy food was motivated by the hypothesis that each food has a unique structure. However, it is not immediately apparent that similar processes could result in similar structures. In the case of the Fried Chicken audio samples, they were obtained from different countries, each with its unique recipe and ingredients. Despite these differences, we treated the Fried Chicken samples as a group to be classified based on their audio features.

Rhythm is defined as a sequence of equal pulses of energy within a given time range. The Beat represents the midpoint of a rhythmic pulse, and in the case of toast, there is only one beat. Although we could identify the amplitude peaks in the audio recordings, we found that these peaks varied depending on the source video. It was impractical to standardize the biting time since the time range of the sound differed between the different types of food. For instance, Fried Chicken had an average duration of 1.3 seconds, Potato Chips had 0.7 seconds, and Toast had 1.2 seconds. The number of samples and average duration of collected sounds are illustrated in Figure 2.6. Since a significant number of audio samples fell outside the one-second standard, it became unfeasible to conduct temporal analysis using neural networks.

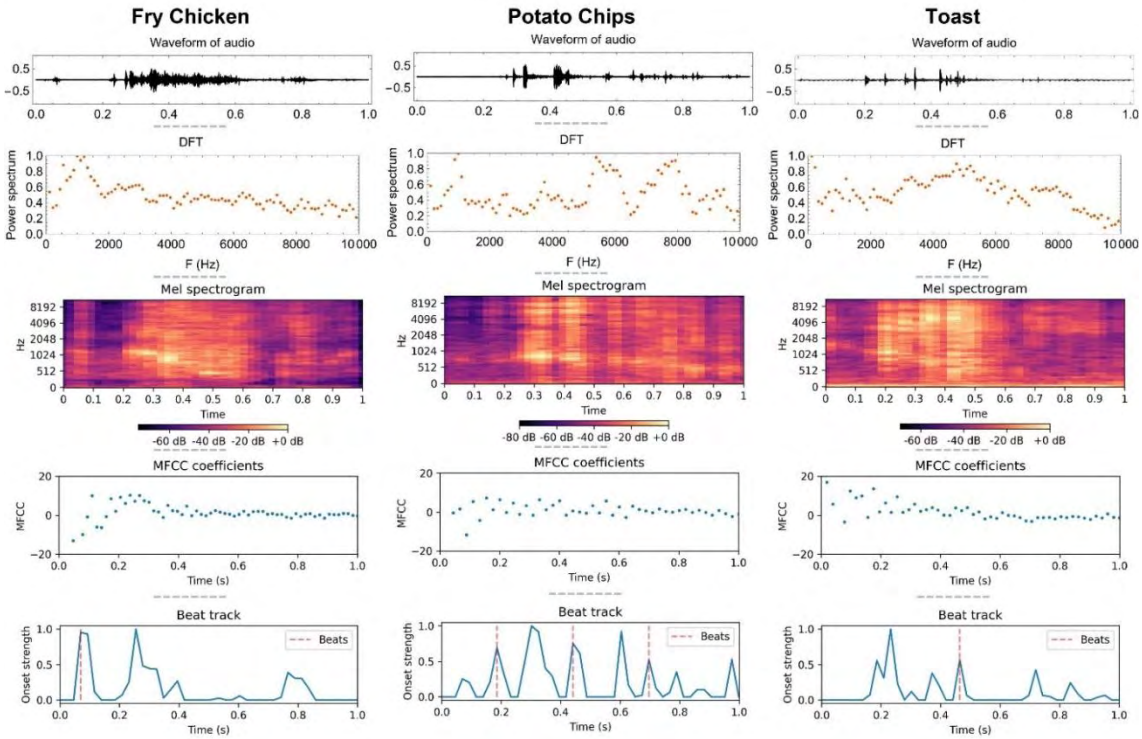
Figure 2.6. Boxplot plot of the variation in the duration of the crispy noise.



The frequency domain provided a more informative visualization of the behavior of the audio recordings over time. Although the amplitude was not readily apparent, the mel-spectrogram revealed potential differences between the audio samples. The brightness of each audio behaved differently, suggesting possible variations in how the crispy noise propagates over time. Figure 2.7 illustrates the behavior of these noises, and although the differences are subtle, they are still detectable. Additionally, the DFT spectrum and MFCC coefficients were given as input examples for the ResNet and MPL models.

The Beat track profiles are utilized to describe the rhythmic pattern of the sounds. In a previous study involving horse audios (ALVES et al., 2021), the three primary features chosen were the MFCCs, the Beat, and the Tempogram. Horses exhibit a distinctive rhythmic pattern in their trot, enabling temporal analysis. The sound produced by the bite generates only one energy pulse, with its peak occurring when both teeth touch. Similar hypotheses were employed in the analysis of the crispy sounds' audio energy.

Figure 2.7. DFT spectrum, Mel spectrogram, MFCC coefficients, and Beat track from the crispy sources: Fried chicken, Potato chips, and Toast.



2.3.2 ResNet Model

The data generated by the discrete Fourier transform (DFT) was utilized as the network inputs for the ResNet model. The length of a DFT array was determined through additional

experiments by varying the moving average filter as 50, 100, 150, and 200 points. The results showed that the highest accuracy was achieved with an input length of 100, which generated an array size of 584x100. In the case of augmented data, the array size increased to 5840x100.

The ResNet achieved an accuracy of 85% when a linear input of 584x100 DFT array values was used, as depicted in Figure 2.8. This outcome provides ample evidence to support the hypothesis that the crispy sound of each food item differed. The phenomenon of overfitting was observed in the result of the DFT network, as evidenced by a validation error spike that occurred after a high number of epochs. This phenomenon is common in machine learning and occurs when the neural network starts to memorize the training data instead of learning to generalize. The overfitting issue persisted regardless of the variation in the number of neurons. To address this issue, the solution proved to be counter-intuitive, as the model required an increase in the number of inputs rather than a reduction.

The augmentation technique multiplied the number of inputs by ten times, obtained 5840 audio inputs, this method solved the tuning problem and boosted the model accuracy to an average of 97% after five pieces of training. Figure 2.9 depicts the improved performance of the ResNet model when using augmented data, as evidenced by the higher model accuracy and lower model loss values.

Figure 2.8. Model accuracy and model loss of the ResNet model for the 584 original data.

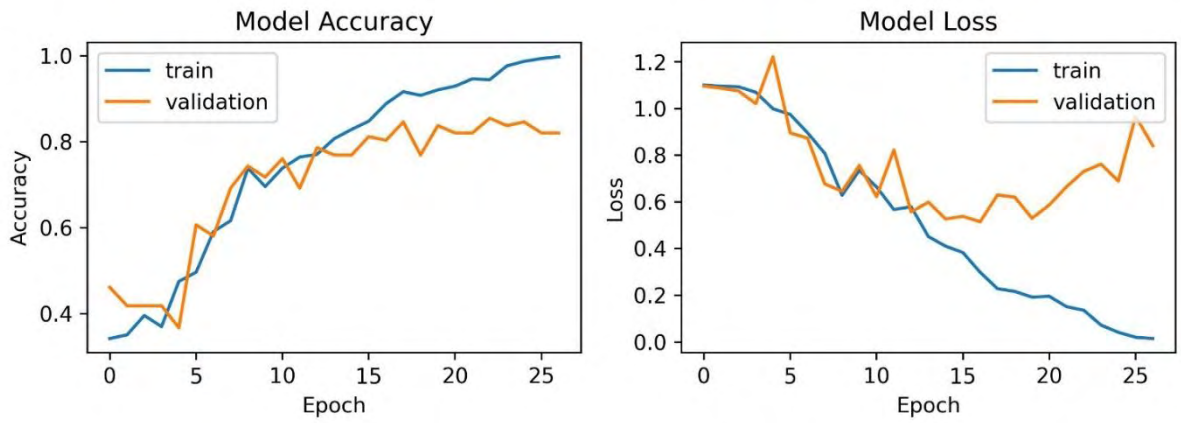
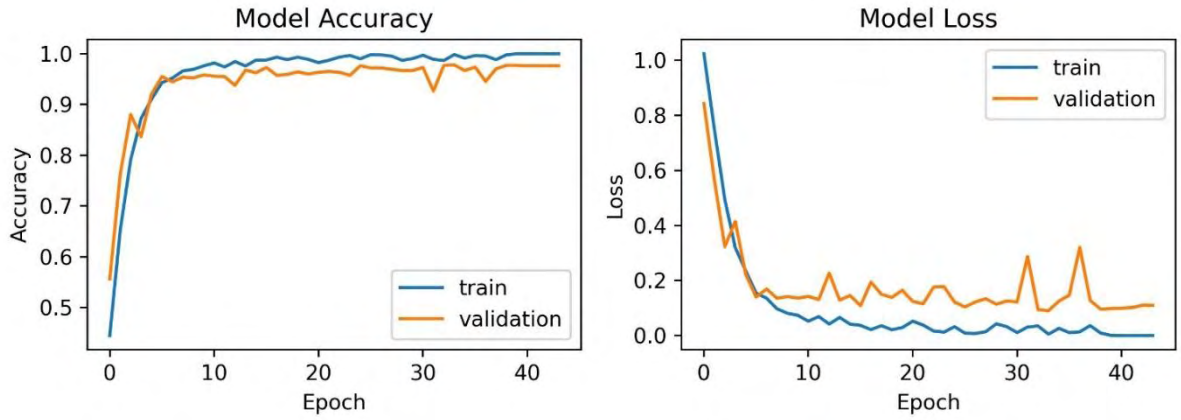


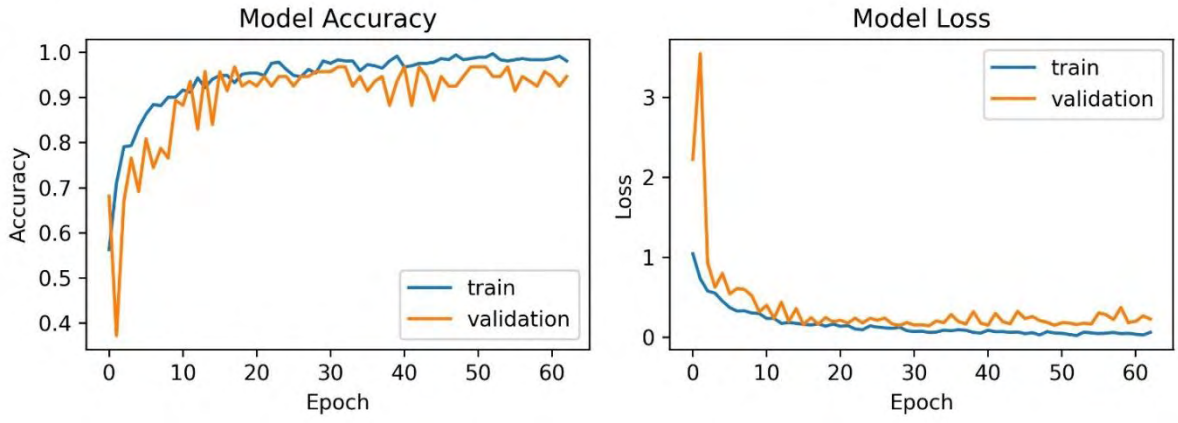
Figure 2.9. Model accuracy and model loss of the ResNet model for the 5840 augmented data.



2.3.3 MLP Model

The convolution holds a relative programming complexity. There was a high processing demand by the Google Collaboratory. The challenge was to find a way that had a lower requirement and was easy to replicate. The MFCC's model attended to the requirements by being a fully connected three-layer network. The simple model reached a peak of 95% accuracy for validation, i.e., it performed satisfactorily, shows in Figure 2.10. Figures 2.8, 2.9 and 2.10 compares the results of the three types of training. All the results arrived at the same end, a high hit rate of the model on what each food type was.

Figure 2.10. Model accuracy and model loss of the MLP model for the 584 original data.



The performance of a neural network also depends on the model's loss, that is, the sum of all the errors the model had in the evaluation stage. The neural network works to minimize this sum of errors, so smaller errors represent better fits. For comparison, the loss of each model

shown in figures 2.8, 2.9 and 2.10 were 0.40, 0.09, and 0.21, respectively. Augmentation brought this performance improvement by proposing an increase in the number of inputs, so the hypothesis is that the other models did not achieve the best performance by not having a larger number of input data.

On the other hand, following the data acquired in a controlled environment, the empirical validation evaluated the model's generalization by using data from a controlled environment. The 50 audios had a different origin compared to the ASMR audios, therefore the model had to guess correctly this completely different audio batch. The DFT models diverged from ideal results, the model without augmentation had a strange behavior. After five training sessions, the evaluation score ranged between 42% and 92%, there wasn't a fixed result for all situations. This could be the overfitting problem; the model didn't identify different patterns between the toast and the fried chicken. It was even worse for the augmentation model with results lower than 20%.

By contrast, the MFCCs outperformed them, after training the model five times, it achieved 100% accuracy in this empirical validation. Even though, the audios had different origins, dividing them into 64 MFCCs proved to be better than analyzing the spectrogram data. The MFCCs split the sound into identities that formed it. For instance, the speech sound is formed by the glottal pulse bypassing the vocal tract. The program split them into two identities: the pulse and the speech timbres. To summarize, any changes in the spectrogram could bring divergent results, but variations didn't affect the MFCCs.

Following the mentioned observations, the best model for a neural network is one that is simpler and demands less manual and computational effort. The MFCC's model meets these requirements very well. It can be stated that for the classification of crispy foods, the MFCC's model is superior to the others presented. It achieved superior results compared to the augmentation model in the empirical validation.

2.3.4 Alternative Neural Networks

In this study, two existing neural network architectures were adapted and evaluated as potential approaches for classifying crispness. The LeNet model was employed, using raw audio, DFT, or Onset Strength signals as inputs. Additionally, the EfficientNetB0 architecture was utilized, with spectrogram images generated by Librosa serving as input for the CONV2D networks (Krizhevsky, Sutskever, & Hinton, 2012). However, both architectures demonstrated inadequate in-sample accuracy in classifying crispness within ASMR videos, highlighting the

ResNet and MLP models as the desired models for this research. The outcomes of each model after 5 training steps are summarized in Table 2.2.

Table 2.2: Comparison of model accuracy among neural network models: ResNet, MLP, EfficientNetB0, and LeNet.

Neural Network model	Type of signal	Size of input Array	Model Accuracy
ResNet with Augmentation	DFT	(5840 x 100 x 1)	97%
MLP	MFCC	(584 x 64)	95%
ResNet	DFT	(584 x 100 x 1)	85%
EfficientNetB0	Spectrogram images	(584 x 224 x 224 x 3)	50%
LeNet with raw audio	Raw audio (2048 Hz)	(584 x 2048 x 1)	48%
LeNet with DFT	DFT	(584 x 100 x 1)	42%
LeNet with Onset Strength	Onset Strength	(584 x 50 x 1)	33%

The complexity level of the tested Artificial Neural Networks (ANNs) can be ordered from higher to lower as follows: ResNet, EfficientNetB0, LeNet, and MLP.

This study reveals that some complex ANN models are not the most reliable for classifying crispness, depending on the input signal, as demonstrated by the EfficientNetB0 model (spectrogram image signal). Additionally, other ANNs with simpler architectures (LeNet and MLP) are not sufficient for classifying the crispness of DFT spectra.

Table 2.2 provides evidence that the models that achieved prediction values higher than 33.3% were able to differentiate the crispness classification between Fried Chicken, Potato Chips, and Toast. However, the LeNet model with the Onset Strength signal showed low accuracy in predicting crispness classification and produced randomized responses.

The in-sample accuracy values between 33.3% and 50% obtained by the LeNet model (DFT and raw audio signals) and the EfficientNetB0 model (Spectrogram images) allowed differentiation of crispness groups, but with low precision and some degree of randomization in the results.

The ResNet model (DFT signal) achieved an in-sample accuracy of 85% and successfully differentiated the crispness of foods. The in-sample accuracy increased to 97% when training the ResNet model with audio data treated by augmentation filters (Timeshift, Reverb, UniformDistribution, and Noise). The use of augmentation expanded the number of inputs by a factor of 10, providing greater differentiation capability.

The MLP model (MFCC) demonstrated a simple ANN architecture and returned high in-sample accuracy (95%) without requiring signal augmentation. The MLP (MFCC) model

proved to be robust, with accurate training and the use of a simple architecture.

Both DFT and MFCC signals are visualized in Figure 2.7. The DFT signal represents the energy distribution across different frequency bands, providing insights into the spectral characteristics of the audio. On the other hand, MFCC analysis evaluates the distribution of formants, which are key acoustic features, over a 1-second time interval.

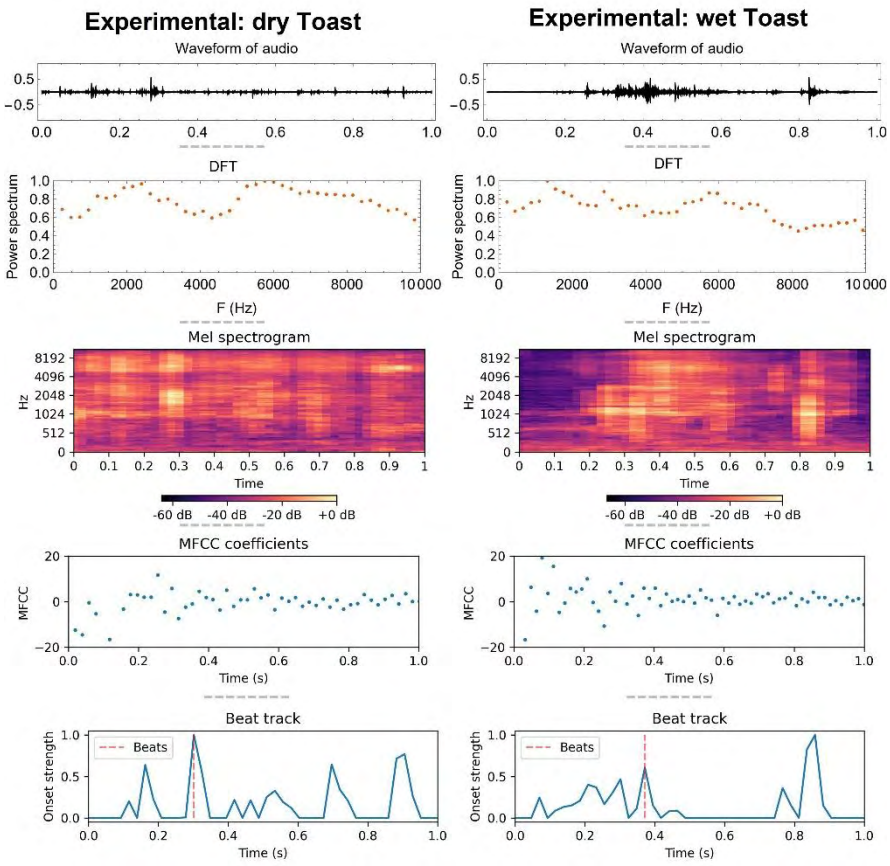
2.3.5 Spectrogram analysis of ASMR mastication audios of fresh toast and toast in milk

The distinction between crispy audios is due to the intrinsic characteristics of each food, and thickness and moisture are examples of properties that could affect spectrum signals.

As observed in the mel spectrograms shown in Figure 2.7, toast has a greater thickness when compared to potato chips. As a consequence, the toast exhibits higher audio energy (dB) within a one-second duration, which can be explained by the brighter area in the mel spectrogram. A similar trend is observed in Figure 2.11 for the spectrograms of fresh toast (dry) and toast previously dipped in milk. The audio energy of toast in milk decreases, indicating a loss of crispness with increasing moisture. This behavior is well-known in shelf-life studies of crispy foods, where toast loses its crispness when stored in a room with higher relative humidity.

When food is fried or dried, water is removed, resulting in a rigid structure filled with air. The audio energy generated by mechanical crushing depends on the air for propagation and the internal structure, which includes porous holes. When the structure is filled with water or liquid, it dampens the sound. A similar effect can be observed with porous materials filled with oil or fat, as illustrated by the mel spectrogram of fried chicken shown in Figure 2.7. The fried chicken has reduced audio energy compared to potato chips or toast.

Figure 2.11. DFT spectrum, Mel spectrogram, MFCC coefficients and Beat track of dry toast, and a toast soaked in milk.



2.3.6 Cross-validation of pre-trained ANNs for fresh toast and toast in milk

The findings from Table 2.3 validate the effectiveness of pre-trained Artificial Neural Networks (ANNs) in evaluating the crispness of fresh toast and toast immersed in milk. The cross-validation experiments incorporated distinct audio sources, including microphone recordings of mechanical crushing experiments (fresh toast) and audio extracted from ASMR videos (toast in milk). Additionally, augmentation audio filtering steps were employed with the original audio recordings to expand the data for both fresh toast and toast in milk samples.

For fresh toast samples (trial 1), the MLP model (MFCC) demonstrated exceptional accuracy (100%) in predicting the crispness of the toast using microphone acquisition. However, the incorporation of audio filter augmentation (trial 4) had an adverse effect on the MFCC signal, leading to a substantial loss in accuracy for the MLP model, reducing it to 35%.

The ResNet (DFT) and ResNet with Augmentation (DFT) models exhibited lower prediction rates (trials 2, 3, 5, and 6). This indicates that the DFT signal has limitations in recognizing the crispness of fresh toast from external sources. Moreover, utilizing audio processed with augmentation filters as input did not improve the prediction performance for

fresh toast samples, resulting in an accuracy below 30%.

When considering ASMR audios of toast in milk, the MLP (MFCC) model once again showed excellent performance, reaching 100% accuracy (trial 7). Therefore, the MFCC signal did not distinguish between fresh toast and toast in milk samples. Additionally, although the MLP model was trained only with dry toast MFCC signals, the accuracy remained high when recognizing the crispness of toast dipped in milk. The MLP (MFCC) model also yielded high predictions for the audios modified by augmentation filters (trial 10), returning an accuracy of 91%. The similarity in the MFCC signal between the crispness of mechanically crushed fresh toast and ASMR toast in milk can be attributed to the similar acoustic properties produced by the crushing process. The resulting sound waves, regardless of the presence of milk, exhibit comparable patterns that the MLP (MFCC) model successfully recognizes and classifies.

For the ASMR toast in milk samples (trials 8, 9, 11, and 12), the ResNet (DFT) model and the ResNet with Augmentation (DFT) model showed comparatively higher prediction rates than the fresh toast samples. This suggests that the audio source from internet ASMR videos provided greater similarity to the pre-training data of the neural networks, including the toast in milk samples. The "imprecise" status was used to mark the trials with accuracy between 50% and 90%. The prediction results for toast in milk when using the ResNet models (DFT) indicate that, although the models were able to make predictions, there is a need for further refinement to achieve more accurate results in classifying external sources. Additionally, it can be inferred that the DFT spectra are more susceptible to variations when the source of external samples changes. This hypothesis directs the application of DFT signals in neural networks that require distinguishing more sensitive differences among pre-trained samples, without external inputs.

The findings demonstrate the effectiveness of pre-trained Artificial Neural Networks (ANNs) in evaluating the crispness of toast and toast in milk. While the MFCC signal proved to be a more robust method for crispness classification, the ResNet models using DFT spectra were found to be more sensitive and less accurate in recognizing external audio sources.

Table 2.3: Cross-validation by pre-trained ANNs using external sources of Fresh Toast and Toast in Milk.

Trial	Sample	Audio source	Pre-trained Neural Network model (Signal)	*Prediction	**Accuracy status
1	Fresh toast	Experimental (microphone)	MLP (MFCC)	100%	High
2	Fresh toast	Experimental (microphone)	ResNet (DFT)	16%	Low
3	Fresh toast	Experimental (microphone)	ResNet with Augmentation (DFT)	19%	Low
4	Fresh toast	Experimental with augmentation filter	MLP (MFCC)	35%	Low
5	Fresh toast	Experimental with augmentation filter	ResNet (DFT)	17%	Low
6	Fresh toast	Experimental with augmentation filter	ResNet with Augmentation (DFT)	30%	Low
7	Toast in milk	ASMR videos (internet)	MLP (MFCC)	100%	High
8	Toast in milk	ASMR videos (internet)	ResNet (DFT)	50%	Imprecise
9	Toast in milk	ASMR videos (internet)	ResNet with Augmentation (DFT)	60%	Imprecise
10	Toast in milk	ASMR videos with augmentation filter	MLP (MFCC)	91%	High
11	Toast in milk	ASMR videos with augmentation filter	ResNet (DFT)	55%	Imprecise
12	Toast in milk	ASMR videos with augmentation filter	ResNet with Augmentation (DFT)	76%	Imprecise

*The values correspond to the mean value (n>10), after 5 training sessions.

** The accuracy status determines if the ANN can accurately predict the crispness of toast using an external input.

2.4 CONCLUSION

This study presents new findings and limitations regarding the use of Artificial Neural Network (ANN) models for analyzing the crispness achieved from audio signals. The method of extracting audio from ASMR videos on internet platforms proves to be a promising approach for pre-training ANN models. Despite considerable variations among random ASMR audios derived from fried chicken, potato chips, and toast sources, both the ResNet model (DFT) with augmented data and the MLP model (MFCC) achieved a high accuracy of over 95% in accurately classifying each crispness sound. To develop a robust neural network model, it was crucial to acquire audio data from various sources and employ different microphones. The resulting model enabled accurate classification of crispy and crunchy foods. The ASMR audio sets were utilized for model training and subsequently validated through experimental tests, demonstrating their potential for expanding the dataset in neural network modeling. It is

important to note that training neural networks to classify the source of crispness must be performed individually, with spectrum data loaded for each food type. This work highlights that each food type has its own crispness recognition, based on the profiles obtained from Discrete Fourier Transform (DFT) and Mel-frequency cepstral coefficients (MFCC). The convolutional residual network, referred to as the ResNet model (DFT), exhibited higher accuracy for classifying crispness "in-sample," especially when augmentation filters were applied to expand the input data. However, the ResNet model demonstrated decreased performance when attempting to classify external inputs of fresh toast or toast in milk. In particular, the ResNet model rejected external inputs but maintained higher in-sample accuracy. On the other hand, the MLP model (MFCC) displayed higher in-sample accuracy and proved to be a more robust tool when compared to the ResNet model, providing a high prediction capability for external inputs, such as fresh toast acquired by a microphone or toast in milk. This research provides a valuable tool for investigating the texture of crispy foods using neural network architectures. It combines data acquisition from high-quality ASMR internet audios and audio from experimental trials involving crushing, enabling a comprehensive analysis of crispy food textures.

CONFLICTS OF INTEREST

On behalf of all authors, the corresponding author states that there is no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors contributed to the study’s conception and design. Material preparation, data collection, and analysis were performed by RZL. Mathematical modeling and audio processing were performed by RZL and GCD. The first draft of the manuscript was written by RZL, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

Funding: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and FAPESP (São Paulo Research Foundation) under grant 2021/05317-9, for their financial support. The authors are grateful for their financial support

REFERENCES

Akimoto, H., Sakurai, N., & Blahovec, J. (2018). A swing arm device for the acoustic measurement of food texture. Journal of Texture Studies, (September 2018), 104–113. https://doi.org/10.1111/jtxs.12381

Alves, A. A. C., Andrietta, L. T., Lopes, R. Z., Bussiman, F. O., Silva, F. F. e, Carvalheiro, R., … Ventura, R. V. (2021). Integrating Audio Signal Processing and Deep Learning Algorithms for Gait Pattern Classification in Brazilian Gaited Horses. Frontiers in Animal Science, 2. https://doi.org/10.3389/fanim.2021.681557

Andreani, P., de Moraes, J. O., Murta, B. H. P., Link, J. V., Tribuzi, G., Laurindo, J. B., … Carciofi, B. A. M. (2020). Spectrum crispness sensory scale correlation with instrumental acoustic high-sampling rate and mechanical analyses. Food Research International, 129, 108886. https://doi.org/10.1016/j.foodres.2019.108886

BUISSON, B., & SILBERZAHN, P. (2010). BLUE OCEAN OR FAST-SECOND INNOVATION? A FOUR-BREAKTHROUGH MODEL TO EXPLAIN SUCCESSFUL MARKET DOMINATION. International Journal of Innovation Management, 14(03), 359–378. https://doi.org/10.1142/S1363919610002684

Chen, L., & Ding, J. (2021). Analysis on Food Crispness Based on Time and Frequency Domain Features of Acoustic Signal. Traitement Du Signal, 38(1), 231–238. https://doi.org/10.18280/ts.380125

Chollet, F. (2017). Deep Learning with Python (1st ed.). USA: Manning Publications Co.

de Moraes, J. O., Andreani, P., Murta, B. H. P., Link, J. V., Tribuzi, G., Laurindo, J. B., … Carciofi, B. A. M. (2022). Mechanical-acoustical measurements to assess the crispness of dehydrated bananas at different water activities. LWT, 154, 112822. https://doi.org/10.1016/j.lwt.2021.112822

Dias-Faceto, L. S., Salvador, A., & Conti-Silva, A. C. (2020). Acoustic settings combination as a sensory crispness indicator of dry crispy food. Journal of Texture Studies, 51(2), 232–241. https://doi.org/10.1111/jtxs.12485

Dorafshan, S., & Azari, H. (2020). Deep learning models for bridge deck evaluation using impact echo. Construction and Building Materials, 263, 120109. https://doi.org/10.1016/j.conbuildmat.2020.120109

Ellis, D. P. W. (2007). Beat Tracking by Dynamic Programming. Journal of New Music Research, 36(1), 51–60. https://doi.org/10.1080/09298210701653344

Gouyo, T., Mestres, C., Maraval, I., Fontez, B., Hofleitner, C., & Bohuon, P. (2020). Assessment of acoustic-mechanical measurements for texture of French fries: Comparison of deep-fat frying and air frying. Food Research International, 131(September 2019), 108947. https://doi.org/10.1016/j.foodres.2019.108947

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition.

Iliassafov, L., & Shimoni, E. (2007). Predicting the sensory crispness of coated turkey breast by its acoustic signature. Food Research International, 40(7), 827–834. https://doi.org/10.1016/j.foodres.2007.01.013

Ji, C., Mudiyanselage, T. B., Gao, Y., & Pan, Y. (2021). A review of infant cry analysis and classification. Eurasip Journal on Audio, Speech, and Music Processing, Vol. 2021. Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1186/s13636-021-00197-5

Kato, S., Ito, R., Wada, N., Kagawa, T., & Yamamoto, M. (2018). Snack Food Texture Estimation by Neural Network. 2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS), 548–553. IEEE. https://doi.org/10.1109/SCIS-ISIS.2018.00097

Kato, S., Wada, N., Ito, R., Shiozaki, T., Nishiyama, Y., & Kagawa, T. (2019a). Snack Texture Estimation System Using a Simple Equipment and Neural Network Model. Future Internet, 11(3). https://doi.org/10.3390/fi11030068

Kato, S., Wada, N., Ito, R., Shiozaki, T., Nishiyama, Y., & Kagawa, T. (2019b). Texture Estimation System of Snacks Using Neural Network Considering Sound and Load. https://doi.org/10.1007/978-3-030-02607-3_5

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, 1097–1105. Red Hook, NY, USA: Curran Associates Inc.

Lai, E. (2003). Converting analog to digital signals and vice versa. In Practical Digital Signal Processing (pp. 14–49). Elsevier. https://doi.org/10.1016/B978-075065798-3/50002-3

Lawless, H. T., & Heymann, H. (2010). Sensory Evaluation of Food - Principles and Practices.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324. https://doi.org/10.1109/5.726791

Levitin, D. J. (2007). This Is Your Brain on Music: The Science of a Human Obsession. New York, NY, USA: Plume Books.

LIU, X., & TAN, J. (1999). ACOUSTIC WAVE ANALYSIS FOR FOOD CRISPNESS EVALUATION. Journal of Texture Studies, 30(4), 397–408. https://doi.org/10.1111/j.1745-4603.1999.tb00227.x

Liu, Y., Cai, M., Zhang, W., Feng, W., Sun, X., Zhang, Y., & Zhou, H. (2021a). Feasibility of non-destructive evaluation for apple crispness based on portable acoustic signal. International Journal of Food Science & Technology, 56(5), 2375–2383. https://doi.org/10.1111/ijfs.14861

Liu, Y., Cai, M., Zhang, W., Feng, W., Sun, X., Zhang, Y., & Zhou, H. (2021b). Original article Feasibility of non-destructive evaluation for apple crispness based on portable acoustic signal. International Journal of Food Science and Technology, 2375–2383. https://doi.org/10.1111/ijfs.14861

Liu, Y., Wu, Q., Huang, J., Zhang, X., Zhu, Y., Zhang, S., … Chen, M. (2021). Comparison of apple firmness prediction models based on non-destructive acoustic signal. International Journal of Food Science & Technology, 56(12), 6443–6450. https://doi.org/10.1111/ijfs.15311

Ma, J. S., Gómez Maureira, M. A., & van Rijn, J. N. (2020). Eating Sound Dataset for 20 Food Types and Sound Classification Using Convolutional Neural Networks. Companion Publication of the 2020 International Conference on Multimodal Interaction, 348–351. New York, NY, USA: ACM. https://doi.org/10.1145/3395035.3425656

O'Shea, N., & Gallagher, E. (2019). Evaluation of novel-extruding ingredients to improve the physicochemical and expansion characteristics of a corn-puffed snack-containing pearled barley. European Food Research and Technology, 245(6), 1293–1305. https://doi.org/10.1007/s00217-019-03260-w

Przybył, K., Duda, A., Koszela, K., & Stangierski, J. (2020). Classification of Dried

Strawberry by the Analysis of. Multidisciplinary Digital Publishing Institute, pp. 1–13.

Sanahuja, S., Fédou, M., & Briesen, H. (2018). Classification of puffed snacks freshness based on crispiness-related mechanical and acoustical properties. Journal of Food Engineering, 226, 53–64. https://doi.org/10.1016/j.jfoodeng.2017.12.013

Spence, C. (2016). Sound: The Forgotten Flavor Sense. In Multisensory Flavor Perception: From Fundamental Neuroscience Through to the Marketplace. Elsevier Ltd. https://doi.org/10.1016/B978-0-08-100350-3.00005-5

Srisawas, W., & Jindal, V. K. (2003). Acoustic testing of snack food crispness using neural networks. Journal of Texture Studies, 34(4), 401–420. https://doi.org/10.1111/j.1745-4603.2003.tb01072.x

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. https://doi.org/10.48550/arXiv.1905.11946

Tunick, M. H., Onwulata, C. I., Thomas, A. E., Phillips, J. G., Mukhopadhyay, S., Sheen, S., … Cooke, P. H. (2013). Critical Evaluation of Crispy and Crunchy Textures: A Review. International Journal of Food Properties, 16(5). https://doi.org/10.1080/10942912.2011.573116

VICKERS, Z. M. (1984). CRISPNESS AND CRUNCHINESS - A DIFFERENCE IN PITCH? Journal of Texture Studies, 15(2). https://doi.org/10.1111/j.1745-4603.1984.tb00375.x

Świetlicka, I., Muszyński, S., & Marzec, A. (2015). Extruded bread classification on the basis of acoustic emission signal with application of artificial neural networks. International Agrophysics, 29(2), 221–229. https://doi.org/10.1515/intag-2015-0022

Zhou, D.-X. (2020). Universality of deep convolutional neural networks. Applied and Computational Harmonic Analysis, 48(2), 787–794. https://doi.org/10.1016/j.acha.2019.06.004

CHAPTER 3

Crispness Quantification of Dry bread and French Fry: A Librosa Approach

3 CRISPNESS QUANTIFICATION OF DRY BREAD AND FRENCH FRY: A LIBROSA APPROACH

Rafael Z. Lopes, Larissa C. Rodrigues, Daniela de Almeida Correa, Gustavo C. Dacanal*

Department of Food Engineering, Faculty of Animal Science and Food Engineering, University of São Paulo, FZEA-USP, 13635-900, Pirassununga, SP, Brazil
* Corresponding author: Gustavo C. Dacanal, E-mail gdacanal@usp.br, Tel/Fax +55 (19) 35654284, ORCID 0000-0002-6061-0981

ABSTRACT

Quality management laboratories often use mechanical and sensory analysis to evaluate crispness. This work proposes an alternative method to represent and analyze crispness using the Librosa package. The dimensionless value Zeta is a correlation between the sound's energy and its intensity over time. The Root Mean Squared Energy (RMSE) calculates the overall magnitude of the audio's energy by compiling its energy over time into a single value, but crispness does not depend only on the energy but on the intensity over time. Therefore, the zeta function (ζ) consists of a multiplication of the RMSE value by the dimensionless peak value. The objective of this work was to create an innovative, fast, and economical method for crispness quantification. The drying process on bread samples simulated an increase in crispness that was measured and compared. The bread was cut into pieces and dried at 60°C for 300 min and the sound crispness was measured every 30 min. For the application on crispness loss, French fries of two different brands were used and the bite sound was captured at 30, 45, and 60 min after frying. It was observed that the more peaks beyond the maximum peak, which indicates breaking, the longer the crispness duration. For the toasted breads, the longer the drying time the more acute were the sounds from 4000 to 6000 Hz, and there is a direct proportionality between the predominant frequency and characteristic sound of the sample. The crispness time increased from 0 to 120 minutes, ranging from 0.2 to 1.2 seconds. The maximum peak consists of breaking while the other peaks can be considered crispy events. The Zeta value for bread had exponential behavior until 90 minutes of drying and after that time the behavior is irregular. For bread, this number varied from 0 to 350 while for French fries the variation was from 2 to 15. Zeta was capable of following the crispness behavior in both experiments pointing out that it is an effective number to quantify crispness.

KEYWORDS:

Dry Bread, French Fry, Quantification, Sound Analysis, Librosa, Crispness.

3.1 INTRODUCTION

Through the years, scientists have changed the way they view crispness, previously it was involved only with mechanical attributes, and now it is subdivided between rheological characteristics and its resulting cracking sound parameters (VICKERS, 2017). Crispness affects the purchase decision and quality perception of the consumers since it indicates product freshness (LAWLESS; HEYMANN, 2010). The sound events of crispy dry food occur due to the structure-breaking sound and its subtle release of air. When applying force with the incisor teeth, energy is retained and dissipated in the form of sound energy during rupture (DIAS-FACETO; SALVADOR; CONTI-SILVA, 2020).

Sensory, mechanical, and acoustical methods can be applied and correlated to measure the food's crispness quality. Sensory methods can provide crispness and sound intensity levels, and mechanical techniques such as the texturometer analysis provide strength and deformation data. Acoustic techniques provide data such as frequency, intensity, and number of sound events or peaks and timing of crispness (TUNICK et al., 2013b). Sensory methods are generally expensive, subjective, and time-consuming, making them unfeasible for routine testing in industries and often the mechanical tests do not correlate with sensory crispness (CHEN; DING, 2021; GOUYO et al., 2020), and the best correlations were obtained between acoustic and mechanical tests (ÇARŞANBA; DUERRSCHMID; SCHLEINING, 2018; PIAZZA; GIOVENZANA, 2015). Studies have been conducted to correlate sound crispness and mechanical crispness by evaluating the sound in the time domain, acoustic signal amplitude, duration, and the number of peaks (AKIMOTO; SAKURAI; BLAHOVEC, 2018; DIAS-FACETO; SALVADOR; CONTI-SILVA, 2020; GOUYO et al., 2020; O'SHEA; GALLAGHER, 2019), but, nowadays, with the use of sound analysis packages it is possible to predict food crispness more quickly, accurately and advantageously (CHEN; DING, 2021; LIU et al., 2021).

French fries and toasted bread are foods that are considered crispy. French fries are easy to prepare and have an attractive flavor, with a crispy dry crust and internal softness. The absorption of oil is desirable, since in the frying process the absorption of oil happens at the same time that moisture is lost, thus the formation of the crispy crust and the formation of pores that are mainly responsible for the crispness occurs (TUNICK et al., 2013b; VAN KOERTEN

et al., 2015). Toasted bread can be considered porous, presenting a mechanical structure that generates good acoustic properties for crispness study, that is, when it suffers deformation with the incisive teeth it releases a great amount of sound energy, but when it presents a humid composition higher than 10% the dry bread practically presents a silent signal (PIAZZA; GIGLI; BENEDETTI, 2008). Thus, due to the high crispness, high noise energy release, and easy processing steps; French fries and toasted bread became the main leads for developing a quantification function using Librosa.

Librosa is a *Python* library that can be used to quickly and easily transform raw audio into the parameters needed to classify and quantify crispness (CHENG et al., 2021; RAGURAMAN; R.; VIJAYAN, 2019). Some works focused on using Neural Networks to classify food products and food fraud using spectrograms, Fast Fourier Transform, Onset Strength, and Mel Frequency Cepstral Coefficients (HUANG et al., 2022; IYMEN et al., 2020; PIAZZA; GIOVENZANA, 2015). However, to make crispness a quantification method easily repeatable by any quality management laboratory, it's needed to transform the right parameters into a final value that can be compared. This work focused on the sound's energy representation and how its intensity unfolds over time. Root Mean Squared Energy computes the overall magnitude of an audio's energy, it compiles its loudness in one value (DWIVEDI; GANGULY; HARAGOPAL, 2023). The crispness representation does not depend only on the loudness, but how loudness unfolds through time. The second parameter used was the number of peaks of the onset strength. Each peak is considered a crispy event. In this work's experiments, more peaks were correlated to higher crispness and less humidity. The sum of all crispy events with their loudness results in the final quantified crispness, Zeta.

Zeta enabled a comparison between the increasing and decreasing of crispness over the processes. While the French Fry experience focused on the delivery and its decreasing crispness over time. The toasted bread experiment was centered on increasing crispness over the drying process time. This work created a crispness quantification method that optimizes quality management in a faster and more affordable way using Zeta as the main lead to evaluate crispness.

3.2 MATERIAL AND METHODS

3.2.1 Material

The organic materials used in the experiments are listed below:

- Pullman Crustless Bread, white bread without its crust. The crust was previously

sliced by the Pullman industries. All bread samples came from the same industrial batch.

- Sadia and McCain French Fries, a Belgian style. They were irregularly sliced, pre-fried potatoes previously stored in a freezer.

3.2.2 Bread sampling

This experiment consists of three replicates using crustless white bread bought in a nearby supermarket. The brand has a decisive factor in this analysis, we want to evaluate the process standardization comparing Zeta over the drying time. The brand chosen came from the Bimbo group (Pullman, Brazil). It came in a squared shape which enabled it to be sliced into four equal-sized pieces. The equipment used in the laboratory tests is a convective oven which is located in the Laboratory of Fluid Dynamics and Characterization of Particulate Systems (LAFLUSP), in the Food Engineering Department of FZEA/USP. The equipment is installed on a bench and it's capable of holding all 80 samples used. Figure 3.1 shows the sample's disposition in the drying machine.

Figure 3.1: Crustless Bread disposition in the convective oven.

The drying experiment consists of a 300-minute batch with eight samples removed from the equipment every 30 minutes. The convective oven's temperature was 60ºC for the batch

(MARCONI TE_037/3, Piracicaba, Brazil), and the bread was disposed into groups of four. They were weighed on an analytical balance. After the convective drying step, the equilibrium moisture ($U_{b.u.}$), on wet basis, was determined by convective drying at 101 °C for 24 hours (Eq. 3.1). From this value, it was possible to determine the dry basis moisture ($U_{b.s.}$) of the samples, by Eq. 3.2. The dimensionless dry basis moisture will be obtained by the Eq. 3.3

$$U_{b.u.} = \frac{m_{\text{water}}}{m_{sample}} \tag{3.1}$$

$$U_{b.s.} = \frac{U_{b.u.}}{1 - U_{b.u.}} \tag{3.2}$$

$$X_{bs}(t) = \frac{U_{b.s.}(t) - U_{eq}}{U_i - U_{eq}} \tag{3.3}$$

3.2.3 French fry sampling

Ingredion Mogi Guaçu hosted the French Fry Analysis. The experiment simulates the delivery of French fries and evaluates the loss of crispness over time. Inside Ingredion's headquarters, the company's Culinology room and Sensory Analysis room simulated the restaurant and the customer's home. The former centered on the preparation of the samples. The French fries were purchased at a nearby market, one from the Sadia brand and one from the McCain brand. The delivery times chosen were zero, 30, 45, and 60 minutes. This reproduces the client receiving the French Fries at their best quality to a sixty-minute wait product. In addition, the sample rate of potatoes that passed the "Crock Tester" was five potatoes at each time.

The preparation began by preheating the vegetable oil from the Soya brand in an electric fryer (Hopkins Electric Deep Fryer with Dual Tank, 3000W, 12 Liters, Stainless Steel, 110 Volts) to 180°C with the aid of a Digital Cooking Thermometer (Facibom). For 5 minutes, 20 potatoes went through the frying process and then separated into groups of five potato sticks each placed in an expanded polystyrene container. The four packages were placed inside an Ifood delivery package made of cardboard.

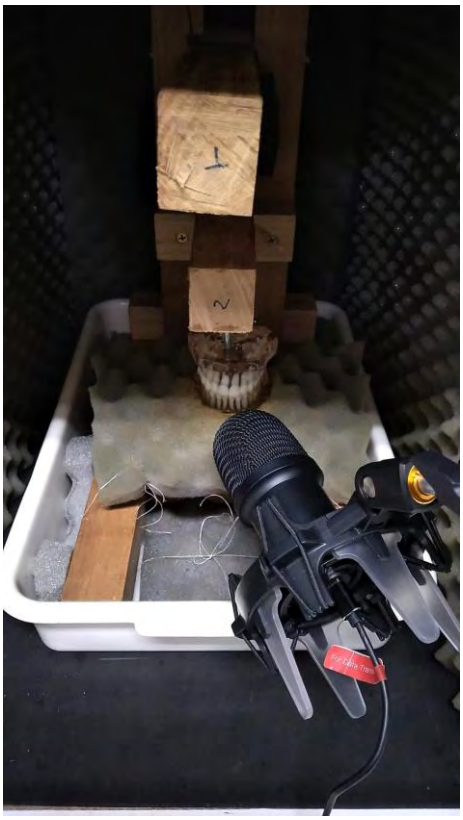
At each of the preset times, the packages were opened and the potatoes were compressed by the dental prosthesis to acquire their crispness. The captured sound followed the same standards of sound treatment and evaluation as the dried bread in Audacity and Python.

3.2.4 Artificial mastication apparatus and acquisition of the crispy noise

The "Crock Tester", the device used in the experiment to capture the sound, consists of a wooden lever that pushes a dental arch simulating the sound of a person biting the toast. Figure 3.2 shows the equipment within the noise suppression box.

The crisp sound was captured by a FIFINE K651 microphone in a mechanical compression test. As the samples were sheared to capture the sound, it's not possible to return them to the oven again.

Figure 3.2: Crock Tester made of a wooden swing arm, a dental prosthesis, and a noise suppression box.



3.2.5 Acquisition of crispy noise from random audios

The crispy audios were acquired from ASMR videos on YouTube from content creators of Japan, Brazil, the United States, Germany, and South Korea. The chosen food were Toast, Fried Chicken, and Potato Chips. The software Shotcut transforms them into 1-second-long audio by cutting the moment people bitted for the first time. These audios formed a small database of 600 audios divided equally among each food type, which is stored in Google Drive.

3.2.6 Audio analysis in Librosa and evaluation of Zeta values

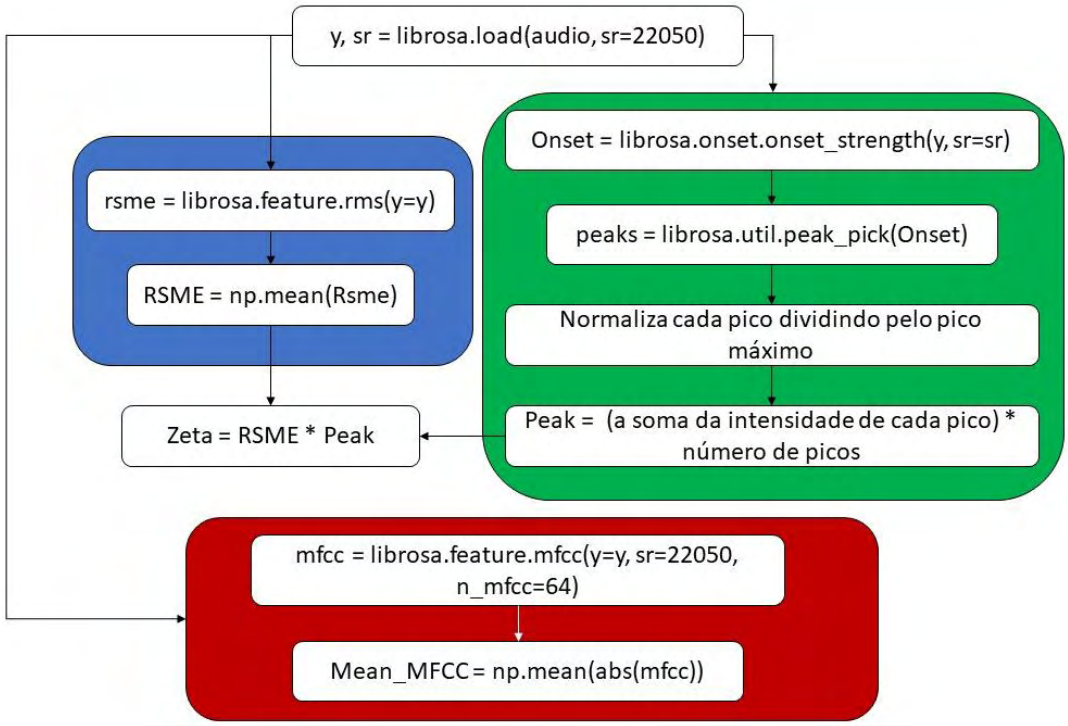
The audio obtained in the Crock Tester was cut into one-second tracks in the Audacity program, and due to the sensitivity of the microphone, Audacity's noise reduction was applied to the samples. Some samples have less than 1 second, hence they were padded with silence as a way to standardize all the audios. After that, in the Python environment, the short-time Fourier Transform was applied to evaluate the sound's energy profile. The first analysis consists of evaluating the following parameters in a 3D normalized spectrogram: time, frequency, and normalized energy. Equations 3.4 and 3.5 define how the normalized energy was defined.

$$E = log10(spectrogram) \quad (3.4)$$

$$En = (E - E_{min})/(E_{max} - E_{min}) \quad (3.5)$$

The flowchart in Figure 3.3 summarizes the audio processing steps in Librosa. The RMSE measures the audio's average energy. The result often ranges between 0.01 and 0.001 for a non-crispy food product, that's we decided to multiply the final value by one hundred.

Figure 3.3: Audio processing flowchart in Librosa.



Similar to the RMSE, the Peak value is obtained by calculating first the onset strength using the *librosa.onset.onset_strength*. These sound impacts are then normalized after the

librosa.util.peak_pick function to create the sound events. The pick peak function has five parameters we need to specify for it to work. The *pre_max* and *post_max* determine the number of samples to consider before and after a max peak, which in this case is the food product rupture. *Pre_avg* and *post_avg* consider an average number of samples before and after a peak; if they are the same, there will be an equal number of peaks before and after the max peak. Usually, crispness consists of the starting "crack" followed by the rupture and its reverberation. Finally, the delta compares the minimum relative height of each peak to its neighbors. In this work, these specifications were all set to one in the function, final values may differ with different entries for this function.

As crispness is correlated to sound, it can be separated into sound events, each one with its intensity but retaining its nature; we decided to call them crispy events. Each peak correlates to the strength of the crispy event, the more picks the longer the duration. Likewise, the bigger the picks, the more intense the crispy events are. Through this analogy, the dimensionless Peak number was obtained by the sum of each normalized peak times the number of the crispy events.

Zeta links the dimensionless energy and loudness to the crispy events. As such, Zeta evaluates the principal characteristics of sound: energy, loudness, and tempo. Equation 3.6 describes how this value is acquired.

$$Zeta = Peak * RMSE \tag{3.6}$$

The Mel Frequency Cepstral Coefficients is the alternative method. As explored in Chapter 2, the coefficients classified crispness by identifying patterns in the 64 Formants. This qualitative method applies a mean to their absolute values and gathers them in a graph. The main purpose is to identify patterns in the graph. These patterns could be additional evidence of a change in the crispness essence. For instance, if the MFCC mean values change drastically, it's possible that the difference in pitch of the sound can be perceived by human hearing.

All data from the three experiments were then gathered and a mean value for the energy, MFCC, Peak intensity, and Zeta were acquired and demonstrated in an Excel graph.

3.3 RESULTS AND DISCUSSION

3.3.1 Dried bread: spectral and temporal analysis

As bread goes through the drying process, a hard crust forms. The structure is also known to give the characteristic sound when fractured; empirically consumers tend to identify what is crisp from the sound so that in the first bite they can already identify if a food emitted a low-intensity sound, is "wilted", or if it is crispy. Studies point out that the cheek acts as a high-frequency filter, making the intonation change (VICKERS, 1984). However, the

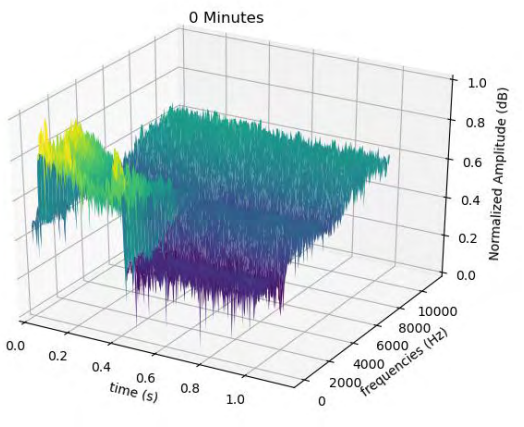
equipment used does not present any kind of filter around the dental prosthesis, which emphasizes that only one medium of propagation was simulated in this study, the airborne medium (VICKERS, 1991).

As the drying time increases, higher frequencies are identified. While at time zero, the predominant frequency range is between 2000 to 4000 Hz with the appearance of frequencies less than 500 Hz. As the drying time passes, the range changes to between 4000 and 6000 Hz, whereas frequencies less than 500 Hz have very little energy identified. This is practical proof that one of the differences is the sound loudness, the more predominant the identified frequency, the more characteristic is the sound. Frequencies lower than 500 Hz are considered bass and they have more relevance from the beginning to the middle of the drying process. On the other hand, due to the appearance of higher frequency ranges, above 4000 thousand, is when the sound became louder and more characteristic. The samples between 30 and 90 minutes indicated a sound similar to a “clench” when going through the crock tester, which reinforces what was indicated in Figure 3.4 as being a lower-pitched sound. Figure 3.4 consists of the data from the first experiment only.

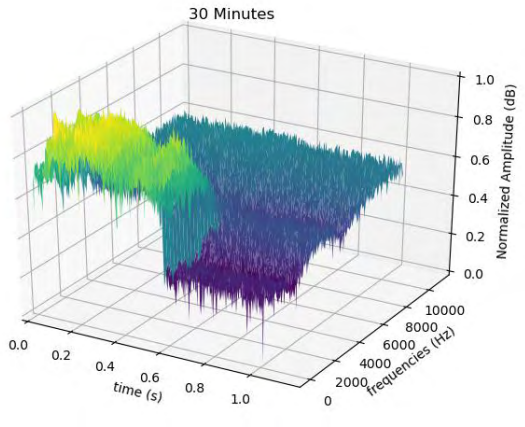
The key factor of this analysis appeared when comparing how the crispy time differs from time to time. They tend to range from 0.2 to 1.2 seconds. The greater the drying time the longer the resulting sound up until a limit. After 120 minutes, the time range didn’t go higher than 1.2 seconds, but the energy increased irregularly.

Figure 3.5 demonstrates the parameters in a 2D perspective. It’s divided between the wave plot, the spectrogram, and the onset peaks. The first one shows the difference in the amplitude of the audio, it starts with values lesser than 0.1 and goes up to 1.0. The sound (a) came from raw bread. It serves as a standard for a zero-crispness bread with almost no amplitude, nor energy, and just one peak. The last one indicates the moment when the food products rupture into two pieces, this situation repeats with each analyzed sound. The highest peak counts as the true peak, therefore it guides how the other peaks will be chosen. Onset strength measures the suddenness of impact.

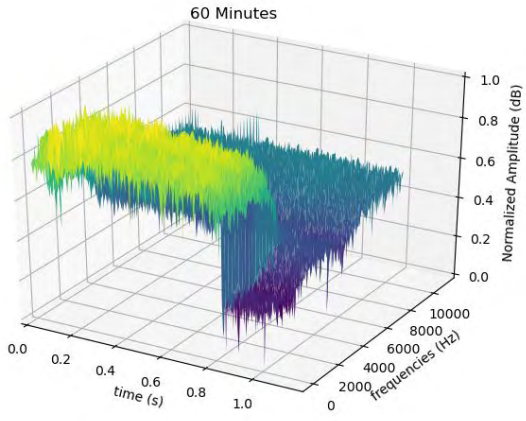
Figure 3.4: 3D spectrograms of toasted bread in (a) 0 minutes, (b) 30 minutes, (c) 60 minutes, (d) 90 minutes, and (e) 120 minutes.



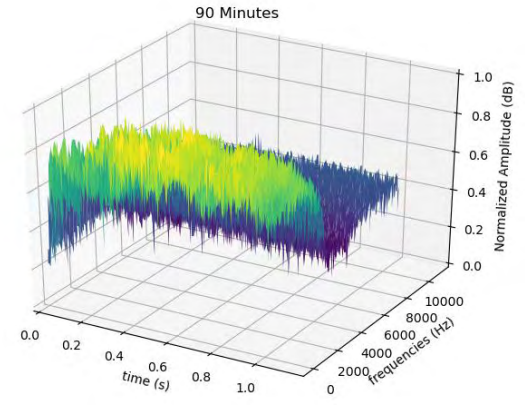
(a)



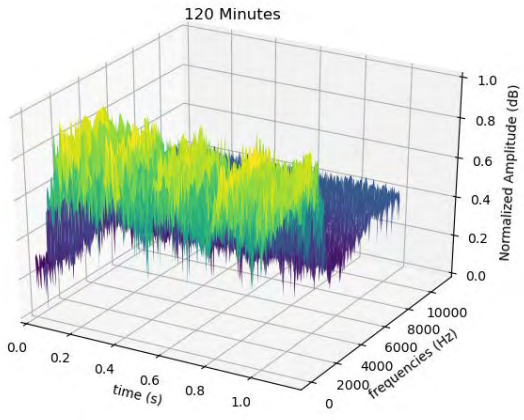
(b)



(c)

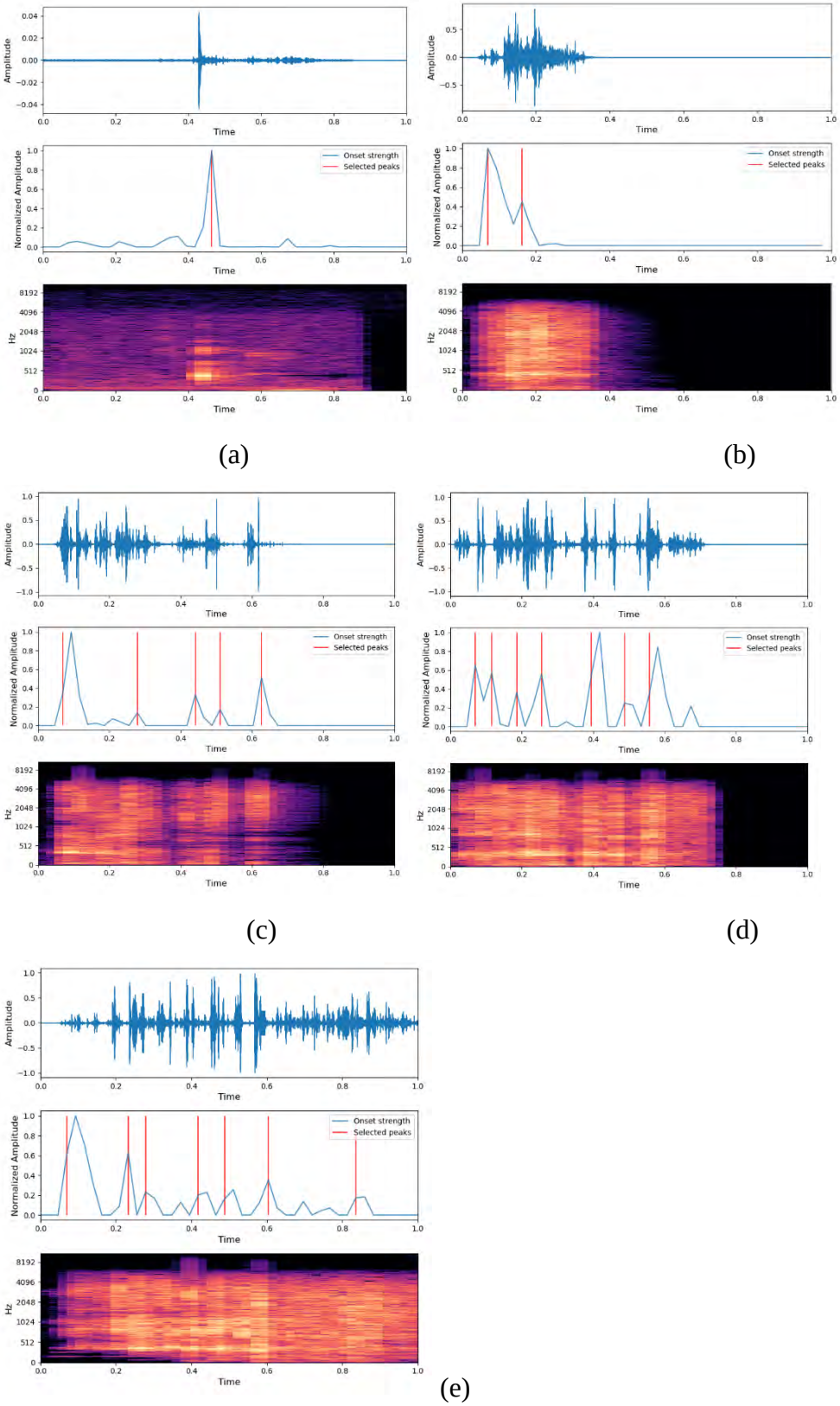


(d)



(e)

Figure 3.5: Comparison of the waveplot, mel spectrogram, and onset peaks of the bread samples in (a) 0 minutes, (b) 30 minutes, (c) 60 minutes, (d) 90 minutes, and (e) 120 minutes.



The main impact is the rupture where the “crack” is noticed, but what are the other peaks? They are the crispy events; they show how crispness resonates through time. These

events may happen before or after the rupture. Its number increases with the drying process up to a limit of eight in one second.

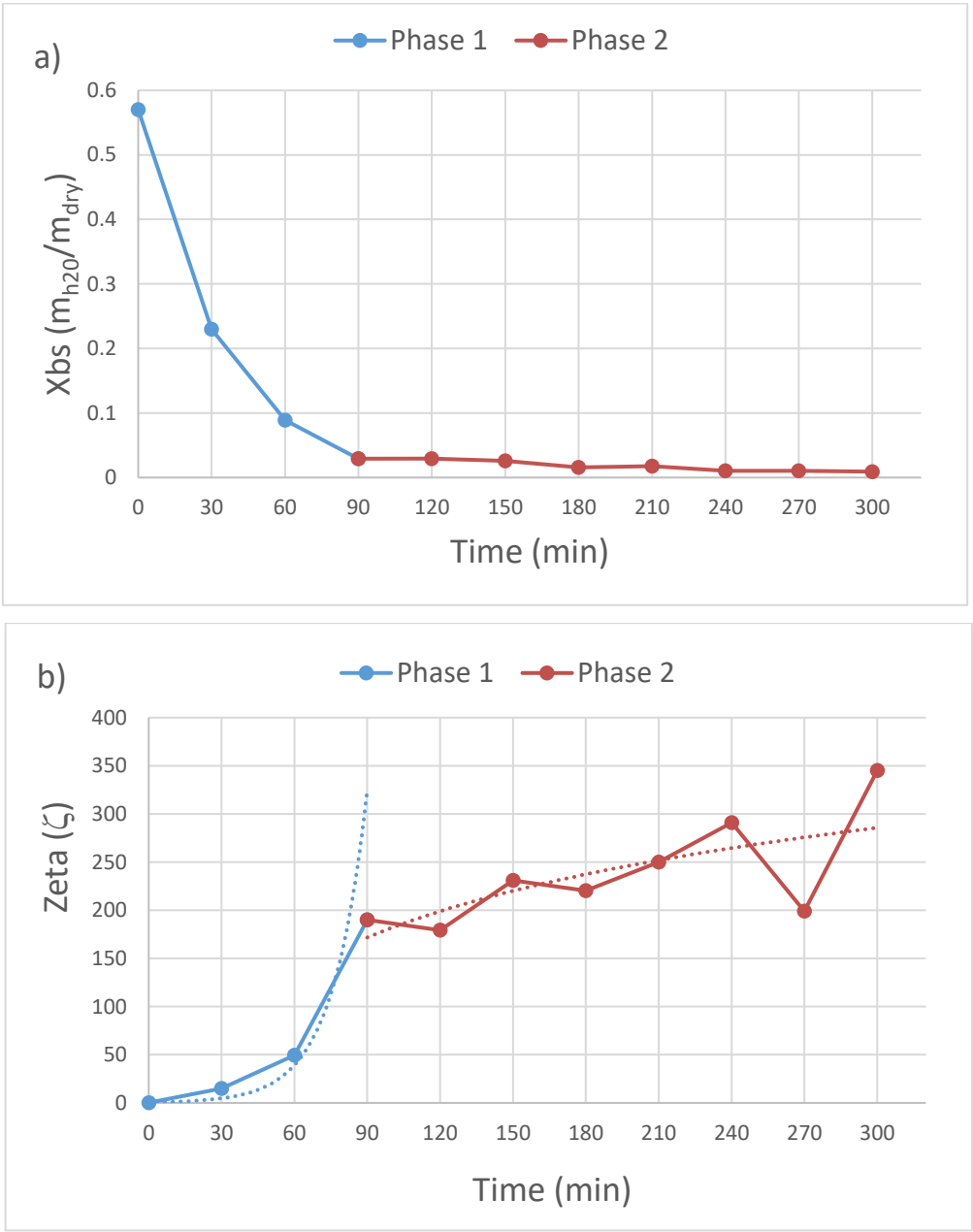
The peaks vary with the audio's time; it was expected that a 300 minutes bread would have the longest crispness time. However, there may be cases where fully dried bread has a crispness time of 0.5 instead of 1 or higher. This affects directly the final value and it isn't controllable. Customer perception is positively influenced by higher crispness lengths. (FILLION; KILCAST, 2002; MALLIKARJUNAN, 2004; VICKERS, 1984, 1985) This time could be standardized by using an intelligent piston that adapts the force input each time. This work used a manual approach to simulate a person biting the toast. He doesn't know how much force it takes to rupture the product, therefore there are weak and strong inputs mixed into the results.

As such, the mean value of these numbers is the resulting perceived crispness. A strong input on dried bread may produce 0.5 or lesser seconds, while a weak input may result in one-second length crispness. As all the samples are standardized to 1 second long, they were padded with silence until it reaches 1 second. This silence didn't change the nature of the audio, but it decreases the number of peaks, and increase the mean energy compared to a sample without the silence.

3.3.2 Dried bread: Zeta compared to the mass fraction of water in the drying process

Figure 3.6 compares how the Zeta behaves through the drying process compared to the decreasing mass fraction of water. These graphs consist of the mean values of the three conducted drying experiments. The graphs were divided into two phases to better compare both situations. Figure 3.6 (b) has two lines that are designed mainly for helping visual interpretation. The first phase consists of an exponential behavior for both (a) and (b). The first 90 minutes is where most of the water content is evaporated, since the structure becomes more rigid, the more energy you need to crack the material increase at the same rate.

Figure 3.6: Zeta compared to the Xbs in two phases: exponential (a) and constant (b)

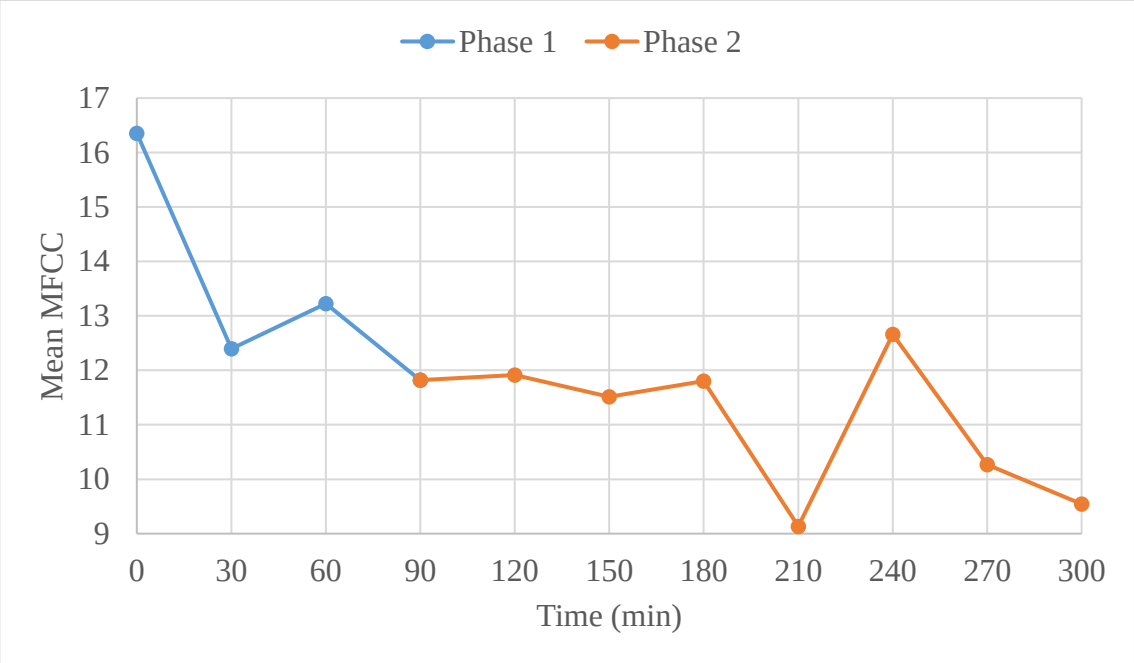


After this time, the behavior became irregular, yet it continues to increase at a slower pace. Although the change in the water content is negligible, the sound became more and more energetic. Zeta varies on the number of crispness events and its energy. In this case, the number of events stayed almost the same, it only raised by one in 240 and 300 minutes. The average energy increased significantly, therefore part of the drying energy stayed within the structure enhancing the crispness. This increase is not high enough to justify an increase in the drying time to get better crispness values. The irregular increase in Phase 2 could be related to differences between each industrial batch. However, a more specific approach should be taken

to validate this possibility.

The MFCC values demonstrate a more qualitative approach to crispness because they point to drastic changes in the Formants. As the MFCCs came from Librosa, higher values indicate a lesser sound intensity, and the reverse is also true. In Figure 3.7, the MFCC mean values behaved differently depending on the experiment's time. In Phase 1, it decreased which is explained by the drying process changing the nature of the sound. In the next Phase, it stagnated in a specific range. This range could be the ideal range where the crispness variation becomes imperceptible by human ears. In the end, there was an irregularity in the values, it's possible that the essence changed again, because when touching the samples they had a sensation of easily crumbling when applying any force.

Figure 3.7: Mean MFCC behavior over time in the dry bread experiment.



3.3.3 French fry: delivery simulation and Zeta behavior

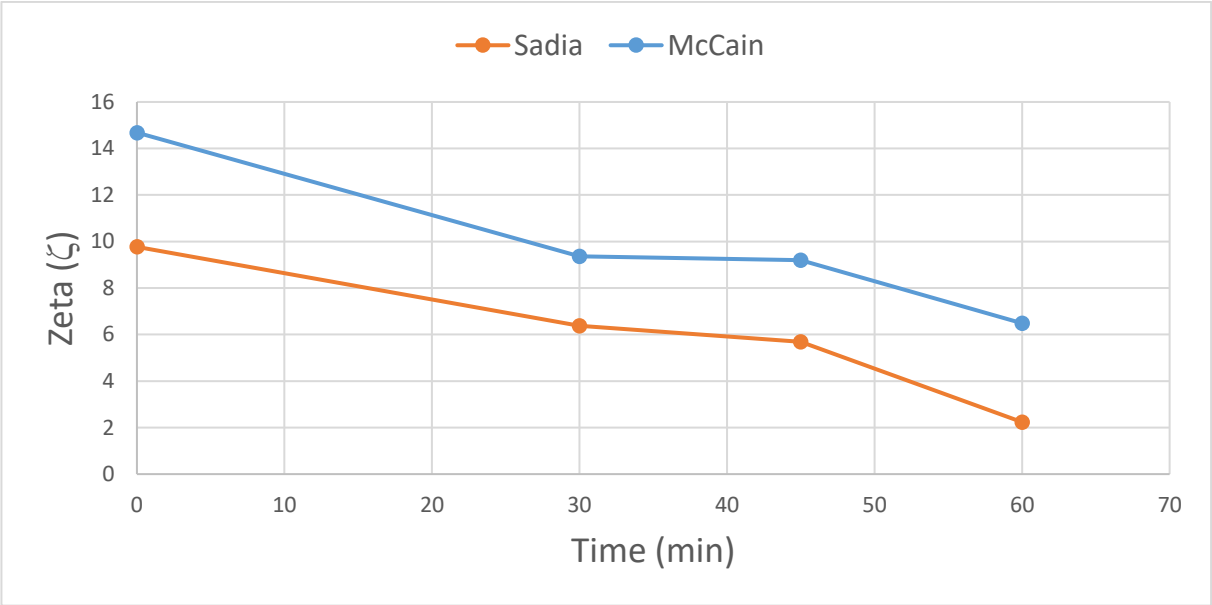
This experiment simulates the delivery of French fries made by a restaurant. This brings a correlation between what your customer expects, a crispy potato, and how intense the sound is when he starts eating it. After capturing more than 100 biting sounds, the ideal Zeta for the French Fry consists of 10 or above. It's the most visible difference compared to the Toast. It's expected to be lower, because of its internal softness. Its thin crust defines the sound's quality. The toast has a thicker crust, therefore more energy and loudness.

The biting test was made using the all Molars. Capturing the sound using the central

incisors only resulted in bad-quality audio with a duration lesser than 0.2 seconds, it's impossible to compare qualities within this shorter and weaker sound. Using the molars permitted a smoother sound that better resembles the sensation of eating French fries.

Coating the French fry brought a significant enhancement compared to the toast without the coating. Figure 3.8 elucidates the difference between them. This study doesn't focus on a deep analysis of coating. Most of the analyses done to assess crispness in Ingredion's Laboratory were sensorial analyses which brought more qualitative than quantitative results. Zeta is a quantitative approach that affects the decision-making for the best coating. The lesser the decrease over time on Zeta the better the coating. This can be a changemaker when choosing which coating will go to the sensory analysis.

Figure 3.8: Zeta behavior over time in the French Fry delivery simulation.



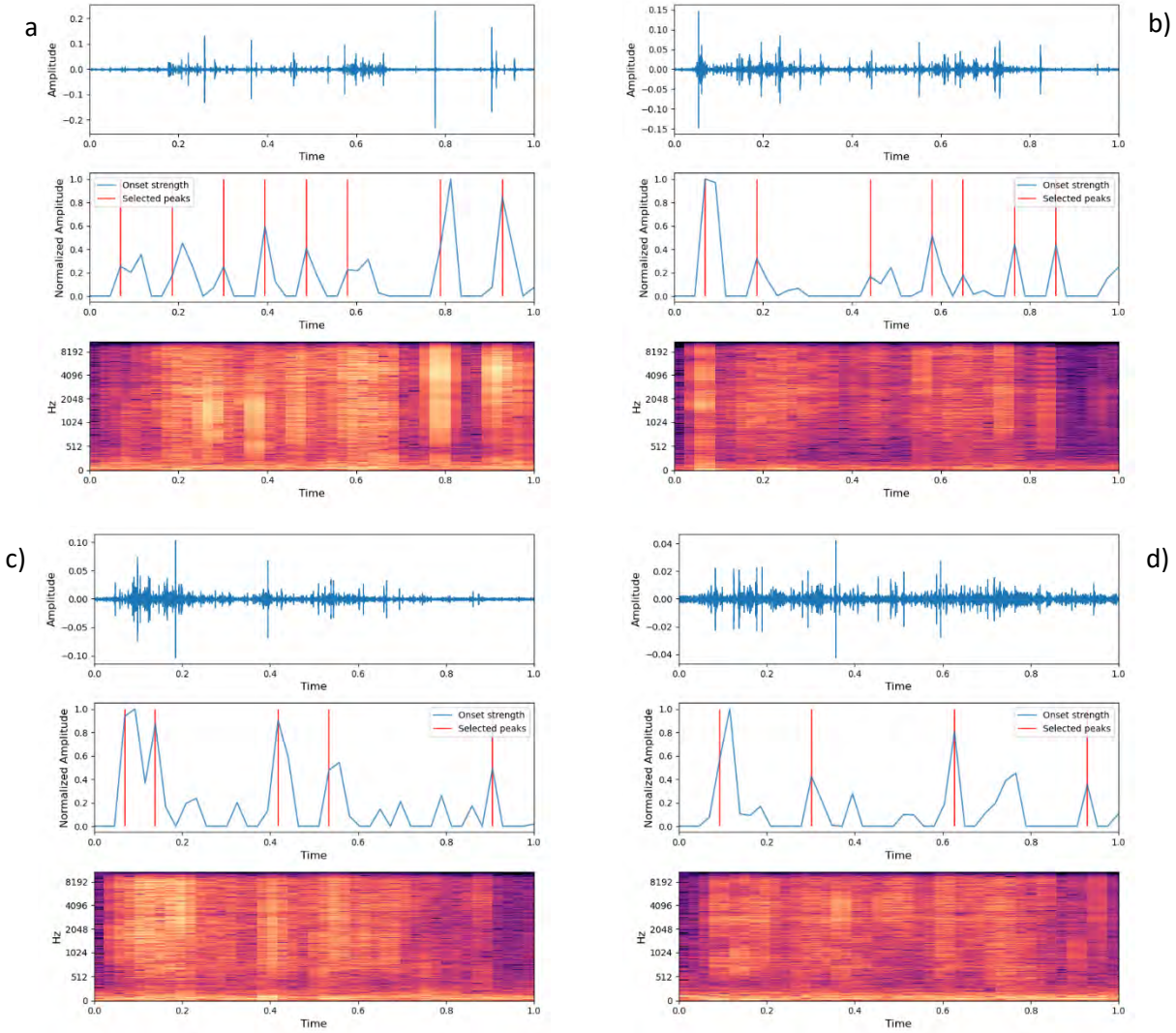
Some of the French fries had "imperfections": a small curvature that resembles the mark of a grid. This happens when you apply the coating and let it be absorbed in a grid for a long time. Although, this small curvature made the sound more energetic increasing the Zeta. This could be a desirable quality instead of an imperfection because the curve is a more rigid structure that intensifies crispness. More investigation on this matter needs to be done in future works to validate this hypothesis

3.3.4 French fry: intensity peaks and spectral analysis

Figure 3.9 extends the comprehension of how Zeta behaved in the experiment. It started

with a lot of energy as shown by the lighter area in (a) and the high number of crispy events. Both values decreased as time passed, but the change in the energy was more evident as the purple zones dominated the spectrogram. Unlike the dried bread, the potato achieved a higher pitched sound bypassing the 10 thousand Hz. This factor influences the perceived crispness by the customer, that's why they evaluate this food product as crispy. (CHANG; VICKERS; TONG, 2018)

Figure 3.9: Comparison of the waveplot, mel spectrogram, and onset peaks of the French Fry samples in (a) 0 minutes, (b) 30 minutes, (c) 45 minutes, and (d) 60 minutes.

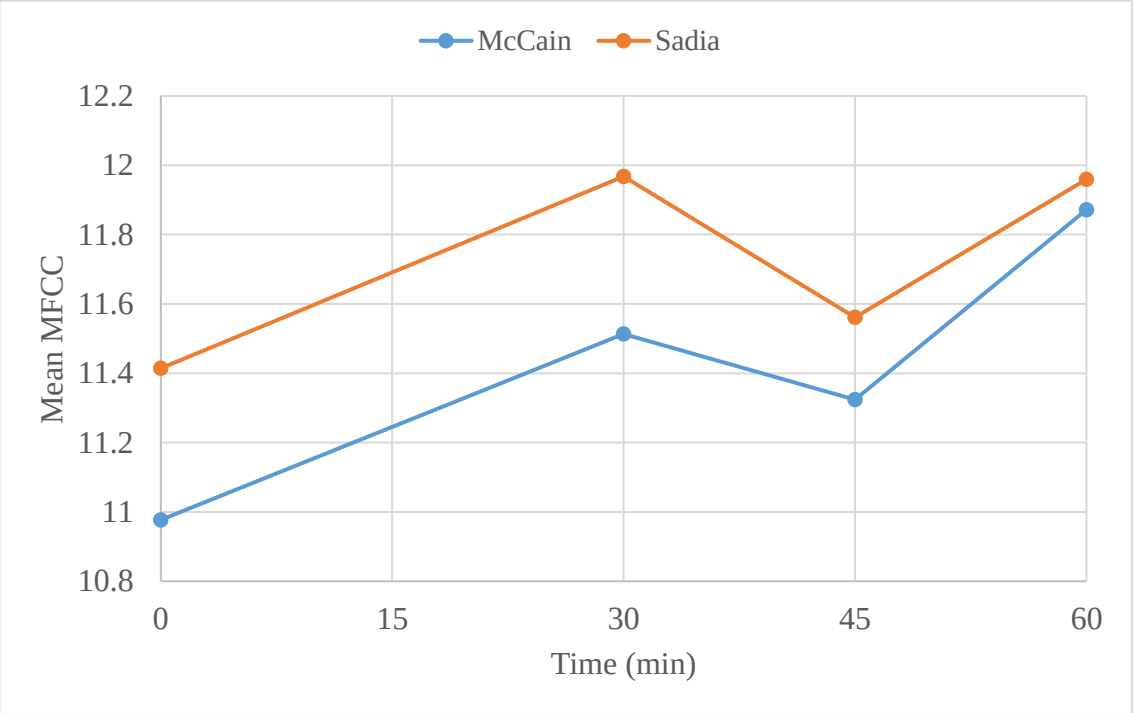


One question to assess is why are these spectrograms so similar, but the Zeta is significantly lower. It's because of the amplitude, the highest achieved value is 0.2. It directly affects the energy and this is shown in the wave plot comparison above. When you compare the size of the wave plots of the dried toast and the French fry, the toast one reaches bigger

amplitude values due to all the factors already discussed. That's why the Zeta is different even though the figures are similar. But they share a similarity, when amplitude gets lower than 0.05, there's no good quality crispness. In brief words, this work suggests that a good quality crispness for the French Fry is correlated with the quality of the energy above the 8 thousand Hz mark and a Zeta superior to 10.

Unlike Figure 3.7, the Mean MFCC in Figure 3.10 didn't have significant variations in the mean value. It demonstrates the sound nature of the crispness maintained over the experiment. However, the Sadia samples appeared to have higher values than the McCain samples. As higher values point to smaller crispness intensity, this is one more piece of evidence that demonstrates the impact of coating in French Fries.

Figure 3.10: Mean MFCC behavior over time in the French Fry experiment.



3.3.5 Evaluation of Zeta in random audios

In Chapter 2, it was possible to classify the crispness of different food materials in random Youtube videos by using neural networks. Table 3.1 elucidate Zeta behavior by calculating a mean and a standard deviation of 200 audio data from different food products. A simple variation such as a different microphone from the one used in this work creates a significant deviation from the numbers shown in Phase 2 of Figure 3.6 (b). Yet, the same problem appeared when analyzing the sound's energy: a slight change in how the sound is

captured causes a significant deviation when comparing the function results (DE MORAES et al., 2022).

Table 3.1: Zeta means and standard deviations of ASMR audios from YouTube of Potato Chips, Fried Chicken, and Toast. Each one has 200 audios.

	Potato Chips	**Fried Chicken**	**Toast**
Mean	135.8	76.7	64.7
Standard Deviation	93.2	81.4	58.8

Controlling the external environment is the key to achieving a more stable analysis. This work suggests an external forcer/piston that applies energy to crack the food products without making any noise. Additionally, crispness most of the time tends to have up to 1 second long, therefore it's relevant to assume this duration on the *librosa.load* function, even though it's lesser than 1 second. Applying silence to these audios doesn't have a significant impact on Zeta.

3.4 CONCLUSION

Zeta behaved similarly to crispness in the experiments. In the drying process, Zeta increased over time as the water mass fraction decreased in comparison. In the simulated delivery of French Fries, Zeta decreased over time as the potatoes lose their crispness. These results indicated that Zeta behaved accordingly to how crispness would behave in those situations. This approach is less complex than building a Neural Network to quantify crispness. Yet, it needs more experimental results to conclude that Zeta is a valid dimensionless number that explains the sound part of the crispness.

The Zeta values on Toasted Bread and French Fry has a disparity of over 20 times between them. This indicates that each crispy food would have its own best crispness range. Toasted Bread's best range in the experiments reached over 200, yet French Fry only achieved next to ten. This work opens paths to new studies to find the best crispness range for each crispy food. This work is the start of a new way to evaluate crispness by sound efficiently. It reached its main objective when the method was tested in a quality management laboratory in Ingredion Mogi Guaçu. It was an easy method for them to evaluate crispness and have data to compare with the sensory analysis.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

ACKNOWLEDGEMENTS

Funding: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and FAPESP (São Paulo Research Foundation) under grant 2021/05317-9. The authors are grateful for their financial support. This study also expresses its gratitude to Ingredion Mogi Guaçu for all the tests and experiments done in their laboratories.

REFERENCES

Akimoto, H., Sakurai, N., Blahovec, J., 2018. A swing-arm device for the acoustic measurement of food texture. J. Texture Stud. 104–113. https://doi.org/10.1111/jtxs.12381

Çarşanba, E., Duerrschmid, K., Schleining, G., 2018. Assessment of acoustic-mechanical measurements for the crispness of wafer products. J. Food Eng. 229, 93–101. https://doi.org/10.1016/j.jfoodeng.2017.11.006

Chang, H.-Y., Vickers, Z.M., Tong, C.B.S., 2018. The use of a combination of instrumental methods to assess change in sensory crispness during storage of a “Honeycrisp” apple breeding family. J. Texture Stud. 49, 228–239. https://doi.org/10.1111/jtxs.12325

Chen, L., Ding, J., 2021. Analysis on Food Crispness Based on Time and Frequency Domain Features of Acoustic Signal. Trait. du Signal 38. https://doi.org/10.18280/ts.380125

Cheng, Y.-H., Chang, P.-C., Nguyen, D.-M., Kuo, C.-N., 2021. Automatic Music Genre Classification Based on CRNN. Eng. Lett. 21, 312–316.

de Moraes, J.O., Andreani, P., Murta, B.H.P., Link, J. V., Tribuzi, G., Laurindo, J.B., Paul, S., Carciofi, B.A.M., 2022. Mechanical-acoustical measurements to assess the crispness of dehydrated bananas at different water activities. LWT 154, 112822. https://doi.org/10.1016/j.lwt.2021.112822

Dias-Faceto, L.S., Salvador, A., Conti-Silva, A.C., 2020. Acoustic settings combination as a sensory crispness indicator of dry crispy food. J. Texture Stud. 51, 232–241. https://doi.org/10.1111/jtxs.12485

Dwivedi, D., Ganguly, A., Haragopal, V.V., 2023. Contrast between simple and complex

classification algorithms, in: Statistical Modeling in Machine Learning. Elsevier, pp. 93–110. https://doi.org/10.1016/B978-0-323-91776-6.00016-6

Fillion, L., Kilcast, D., 2002. Consumer perception of crispness and crunchiness in fruits and vegetables. Food Qual. Prefer. 13, 23–29. https://doi.org/10.1016/S0950-3293(01)00053-2

Gouyo, T., Mestres, C., Maraval, I., Fontez, B., Hofleitner, C., Bohuon, P., 2020. Assessment of acoustic-mechanical measurements for texture of French fries: Comparison of deep-fat frying and air frying. Food Res. Int. 131, 108947. https://doi.org/10.1016/j.foodres.2019.108947

Huang, T.-W., Bhat, S.A., Huang, N.-F., Chang, C.-Y., Chan, P.-C., Elepano, A.R., 2022. Artificial Intelligence-Based Real-Time Pineapple Quality Classification Using Acoustic Spectroscopy. Agriculture 12, 129. https://doi.org/10.3390/agriculture12020129

Iymen, G., Tanriver, G., Hayirlioglu, Y.Z., Ergen, O., 2020. Artificial intelligence-based identification of butter variations as a model study for detecting food adulteration. Innov. Food Sci. Emerg. Technol. 66, 102527. https://doi.org/10.1016/j.ifset.2020.102527

Lawless, H.T., Heymann, H., 2010. Sensory Evaluation of Food - Principles and Practices.

Liu, Y., Cai, M., Zhang, W., Feng, W., Sun, X., Zhang, Y., Zhou, H., 2021. Original article Feasibility of non-destructive evaluation for apple crispness based on portable acoustic signal. Int. J. Food Sci. Technol. 2375–2383. https://doi.org/10.1111/ijfs.14861

Mallikarjunan, P., 2004. Understanding and measuring consumer perceptions of crispness, in: Texture in Food. Elsevier, pp. 82–105. https://doi.org/10.1533/978185538362.1.82

O'Shea, N., Gallagher, E., 2019. Evaluation of novel-extruding ingredients to improve the physicochemical and expansion characteristics of a corn-puffed snack-containing pearled barley. Eur. Food Res. Technol. 245, 1293–1305. https://doi.org/10.1007/s00217-019-03260-w

Piazza, L., Gigli, J., Benedetti, S., 2008. Study of structure and flavour release relationships in low moisture bakery products by means of the acoustic e mechanical combined technique and the electronic nose 48, 413–419. https://doi.org/10.1016/j.jcs.2007.09.016

Piazza, L., Giovenzana, V., 2015. Instrumental acoustic-mechanical measures of crispness in apples. Food Res. Int. 69, 209–215. https://doi.org/10.1016/j.foodres.2014.12.041

Raguraman, P., R., M., Vijayan, M., 2019. LibROSA Based Assessment Tool for Music Information Retrieval Systems, in: 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). IEEE. https://doi.org/10.1109/MIPR.2019.00027

Tunick, M.H., Onwulata, C.I., Thomas, A.E., Phillips, J.G., Mukhopadhyay, S., Sheen, S., Liu, C.-K., Latona, N., Pimentel, M.R., Cooke, P.H., 2013. Critical Evaluation of Crispy and Crunchy Textures: A Review. Int. J. Food Prop. 16, 949–963. https://doi.org/10.1080/10942912.2011.573116

van Koerten, K.N., Schutyser, M.A.I., Somsen, D., Boom, R.M., 2015. Crust morphology and crispness development during deep-fat frying of potato. Food Res. Int. 78, 336–342. https://doi.org/10.1016/j.foodres.2015.09.022

VICKERS, Z., 1991. SOUND PERCEPTION AND FOOD QUALITY. J. Food Qual. 14, 87–96. https://doi.org/10.1111/j.1745-4557.1991.tb00049.x

Vickers, Z.M., 2017. Crispness and Crunchiness— Textural Attributes with Auditory Components, in: Food Texture. Routledge, pp. 145–166. https://doi.org/10.1201/9780203755600-6

VICKERS, Z.M., 1985. THE RELATIONSHIPS OF PITCH, LOUDNESS AND EATING TECHNIQUE TO JUDGMENTS OF THE CRISPNESS AND CRUNCHINESS OF FOOD SOUNDS. J. Texture Stud. 16, 85–95. https://doi.org/10.1111/j.1745-4603.1985.tb00681.x

VICKERS, Z.M., 1984. CRISPNESS AND CRUNCHINESS - A DIFFERENCE IN PITCH? J. Texture Stud. 15. https://doi.org/10.1111/j.1745-4603.1984.tb00375.x

CHAPTER 4

General Conclusion and Final Remarks

4 General Conclusion and Final Remarks

Crispness behavior was analyzed through two practical methods. The neural network method pointed out that crispness is different depending on the food type. The product's structure impacts the sound's loudness and intensity. The energy spectrogram varied depending on the product thickness, water mass fraction, and how the crispy crust formed. Toast is thicker than a Potato Chip, Fried Chicken has a higher water mass fraction than toast. The resulting sound profile could be successfully classified by MFCC's and FFT models. MFCC's could even be used to identify variations in the crispness patterns. It was able to identify all the 50 Toast samples collected in a controlled environment, even if the microphone used was different from the ASMR videos.

The quantification method demonstrated that it's possible to reach a dimensionless number that behaves accordingly to expectations. However, it needs a controlled environment to be viable. The variables such as the microphone and the force input to bite the crispy food interfere substantially with the final value. Different microphones impacted the FFT classification method while the crispness time range dictated in the number of onset peaks. These variables resulted in the irregularities in Phase 2 in Figure 3.6.

The MFCC as a qualitative method to perceive the sound's essence and verify changes in its nature proved to be a good tool for evaluating crispness. As MFCC's doesn't vary with the crispness length, 0.2 and 1-second-long crispness could be evaluated as having the same nature.

This work is the first step in the creation of a Software capable of evaluating crispness behavior precisely. Although it has some irregularities, the methods can be refined and polished until perfection with new approaches. This Master's thesis successfully opened new research perspectives to the crispness that can be further explored in the next years.