Bone Conduction-like Acoustic Sensor System for Evaluating Crispness

Yasutoshi Makino[1], Nobutaka Ono[1], Shigeru Ando[1],
Fumihiko Sano[2], and Shigeru Toba[2]
1 Graduate School of Information Science and Technology, University of Tokyo
7–3–1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
2 Ajinomoto Frozen Foods co.,inc.
1210–5 Oaza, Yoshida, Oizumi-machi, Ora-gun, Gunma, 370-0523, Japan
Phone: +81–3–5841–6927, Email:`yasutoc@alab.t.u-tokyo.ac.jp`

Abstract: In this paper, we propose a novel acoustic sensor system to evaluate *crispness* by focusing localization of biting sounds and its spatial spread. With mimicking bone conduction mechanism in human, our system consists of 1) acrylic teeth to bite food, 2) an acrylic bone to propagate sounds (vibrations) generated by food, 3) two acoustic emission (AE) sensors to obtain the vibrations. From left and right sensor outputs, this system localizes sound sources in biting and obtains their spatial spread. Although such features have never been used to estimate texture in food, we show they clearly classify four characteristic crispy foods by experiments.

Keywords: crispness, bone conduction, spatial spread, food, biomimicry

1. Introduction

Since *Crispness* is one of the essential properties to many familiar foods like croquette, serial, potato chips, and so on, evaluation of crispness is of large interest in food science field. Today, sensory test has been widely used for it, however, the method is not objective and its results are influenced by external environments such as temperature, humidity, appetite of the subjects.

Since Vickers[2, 3] showed that auditory information is as important as taste or tactile one, many research have attempted to evaluate crispness by acoustic properties generated by foods, such as amplitudes, height of peaks, number of bursts[6], frequency characteristics[4], fractal dimension[8]. Most of them are based on monaural recorded sounds.

In this paper, we focus on localization of sounds and its spatial spread, which are much important sensation in binaural hearing. In order to obtain them, we mimic bone conduction mechanism in human because it is well known that bone conductive sounds contribute to auditory sensation as well as air conductive sounds in mastication[1, 4]. So, our sensor system consists of 1) acrylic teeth, 2) an acrylic bone, 3) two acoustic emission (AE) sensors as ears, and from binaural sensor outputs, this system localizes sound sources in biting and obtains their spatial spread. Although such features have never been used to estimate texture in food, we show they clearly classify four characteristic crispy foods by experiments.

2. Sensor System

2.1 Bone Conduction-like Acoustic Sensor System

Fig.1 shows our sensor system. The vibrations of biting sounds are detected from teeth and propagate through jawbone then ears sense the vibrations. Using the time lag and the differences of strength of sounds between both ears, we can calculate values related with spatial spread.

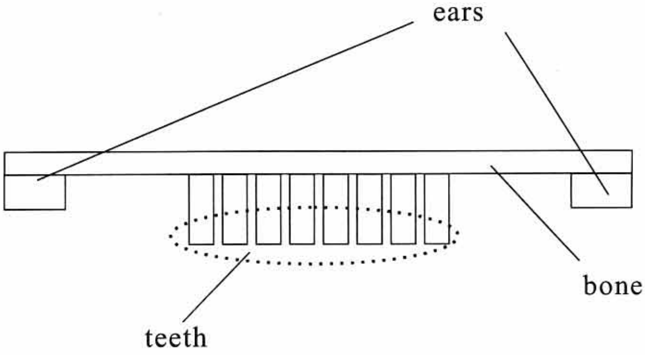


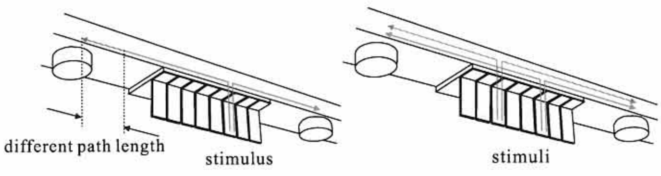
Fig. 1: The concept of bone conduction-like sensor

2.2 Sensing Spatial Spread of Biting Sounds

In our system, spatial information in mastication is detected by the Spatio-Temporal Gradient methods[9]. Here, we describe its principle simply.

Assume that food produces a vibration around only one tooth (Fig.2(a)). Then, the vibration propagates from the tooth, through the bone and arrived at left and right sensors. Since the different path length yields time lag and different attenuation, the waveform observed at the sensors are written by

$$f_R(t) \cong (1+\xi)f(t+\tau), \quad (1)$$
$$f_L(t) \cong (1-\xi)f(t-\tau), \quad (2)$$



(a) (b)

Fig. 2: The concept of sound source localization (a) single stimulus and its different path length. (b) multi stimuli

where τ and ξ denotes time difference and intensity difference, respectively. By the first order approximation, these equations deform to

$$\begin{aligned} f_R(t) &\cong f(t) + \tau\dot{f}(t) + \xi f(t), \quad (3)\\ f_L(t) &\cong f(t) - \tau\dot{f}(t) - \xi f(t). \quad (4)\end{aligned}$$

The sum and difference of both side are

$$\begin{aligned} f_+(t) &\equiv f_R(t) + f_L(t) \cong 2f(t), \quad (5)\\ f_-(t) &\equiv f_R(t) - f_L(t) \cong 2\tau\dot{f}(t) + 2\xi f(t). \quad (6)\end{aligned}$$

Then, replacing $f(t)$ with $f_+(t)$, we obtain

$$f_- \cong \tau\dot{f}_+ + \xi f_+. \quad (7)$$

By minimizing the square of errors as

$$J_{\text{res}} = \int_\Gamma |f_- - \xi f_+ - \tau\dot{f}_+|^2 dt, \quad (8)$$

we obtain the optimal estimation of τ and ξ, and localize the sound source in biting. Moreover, the minimum value of J_{res} represents suitableness for the model, (1), (2).

When two or more sounds are generated at different teeth, independently (Fig.2(b)), sensor outputs are not satisfied with the model, (1) and (2). Then, the minimum of J_{res} takes a large value. Therefore, the spatial spread of sounds (hardness to localization) is estimated by

$$Q_1 = \frac{J_{\text{res}}}{\int_\Gamma |f_-|^2 dt}. \quad (9)$$

2.3 Fabrication

To reduce standing waves on the bone, we install shock absorber between the bone part and aluminum support rod. Rubber sheet between teeth and the bone makes it easy for teeth to vibrate with foods. Their sizes are shown in Table.1

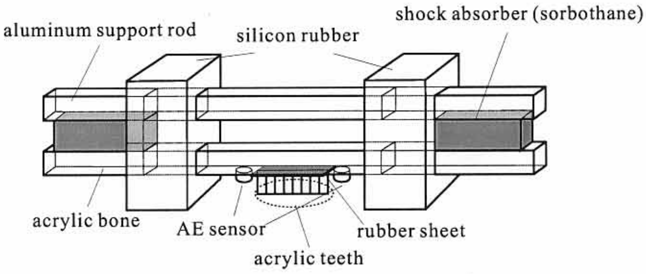


Fig. 3: Sensor system

region	materials	size
jawbone	acrylic	15mm× 15mm×500mm
teeth	acrylic	5mm× 15mm× 1mm
contact	rubber sheet	47mm× 15mm× 1mm
support unit	aluminum	5mm× 5mm×600mm
	silicon rubber	53mm×100mm× 70mm

Table. 1: Size of materials

Fig.4 shows our sensor unit. This unit consists of four of the followings, bone conduction-like sensor system, automatic stage, amplifier and A/D translation machine. An automatic stage moves up and down, then food is crushed and vibration is detected by bone conduction-like sensor system. The signal is amplified about three hundred times and A/D translation is carried out before being taken in PC.
Fig.5 is photo of our experiment system taken from the front and Fig.6 is dental enlargement.

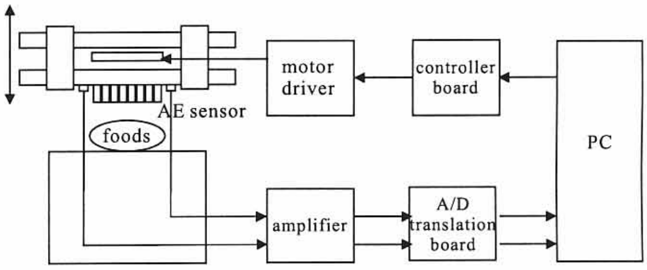


Fig. 4: Block diagram

3. Feature Extraction

3.1 Frequency Decomposition

The signals we had detected were decomposed into four frequency ranges in order to analyze easily. In fact, Frequency decomposition is performed in our cochlea. We used following formula as a filter.

$$H(\omega) = e^{-\frac{(\omega-\omega_0)^2}{2\Delta\omega^2}} \quad (10)$$

400, 800, 1200 and 1600 (Hz) were used as the $\omega_0/2\pi$ and 400(Hz) was applied as the $\Delta\omega/2\pi$ (Fig.7).

3.2 Feature values

To evaluate crispness, we calculate following four kinds of features.

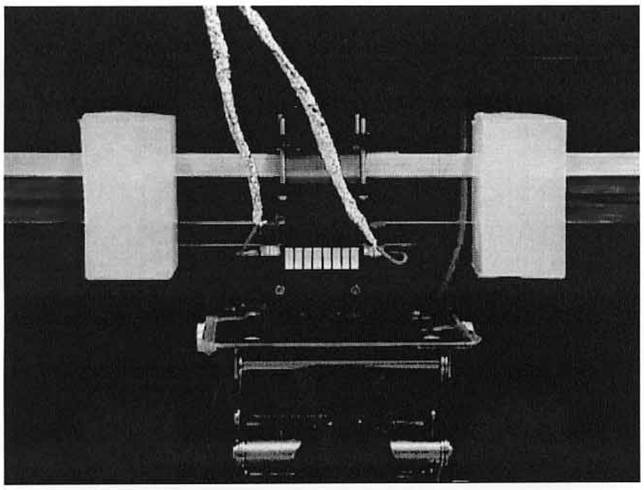


Fig. 5: Experiment system

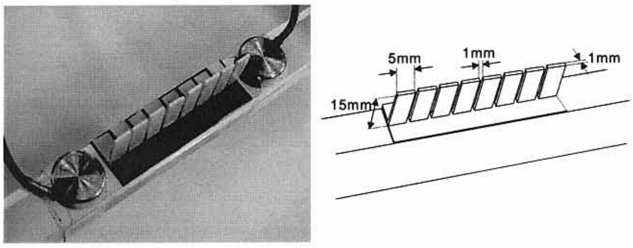


Fig. 6: Dental enlargement

1) average of Q_1:

This value is shown as $\overline{Q_1} = \frac{\sum_i Q_{1_i}}{N}$. This means whether approximation is appropriate or not. Crispy food may present high value.

2) differential variance of τ:

$\Delta\tau = \sqrt{\frac{\sum_i (\tau_i - \tau_{i-1})^2}{N}}$ This value shows spread of position of destructions. When single stimulus is applied this value decreases because the time lag τ is constant.

3) average frequency of signal:

We use $\sqrt{\frac{\int |\dot{f}(t)|^2 dt}{\int |f(t)|^2}}$ to calculate th average frequency of signal. This really means average of ω^2 not ω, however, if we get $A\sin(\omega t + \phi)$, the differentiation of this single frequency signal is given $A\omega\cos(\omega t + \phi)$ and this formula is filled.

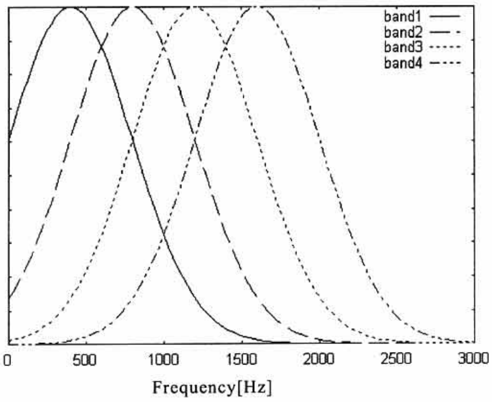


Fig. 7: Using band range

4) average frequency of power:

We get average frequency of power using $\sqrt{\frac{\int |\dot{S}(t)|^2 dt}{\int |S(t)|^2}}$. If the destruction is caused with short duration, this value increases and long duration decreases.

1), 2) and 3) is calculated in each band and 4) is carried out for the original signal.

4. Experiments and Results

4.1 Principal Component Analysis for Crispy Foods

The following four foods were prepared as crispy ones.

(a) cracotte

(b) apple

(c) potato chip

(d) pocky

In order to bite foods, we drove teeth only downward with the speed 5cm/s. After detection of biting sounds, we calculate the thirteen features described the previous subsection. Then we performed the principal component analysis(PCA) by them. Fig.8 shows the result, in which each sample is plotted in the PC1-PC2 space because the total power of PC1 and PC2 achieve 63% to the whole one. It shows that the four foods are clearly classified in the feature space. PC1 and PC2 nearly corresponds to localized sensation and variation of sound sources, respectively.

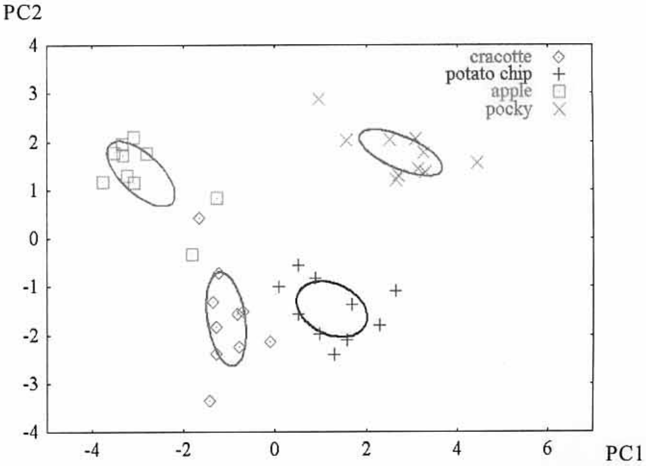


Fig. 8: The result of PCA with four crispy foods. The ellipse is drawn by variation respectively

4.2 Evaluation of the Feature Space

We prepared two foods to verify whether our feeling of texture is matched with the result in section 4.1. Wafer and Ebisenbei were examined by the same way as previous section. Fig.9 shows the result of wafer. They plotted between the cracotte and the potato chip. Compare with our feeling, the points they were plotted are appropriate positions. Ebisenbei also shows same result.

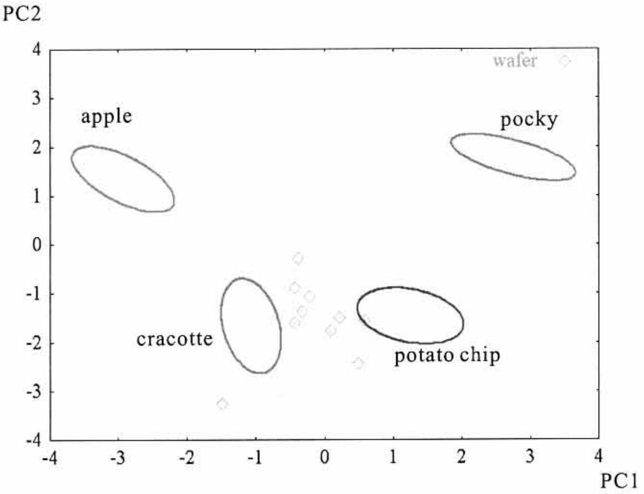


Fig. 9: The result of wafer plotted on feature space

4.3 Two Croquettes

We want to know the differences of the same food except process. Hence, delicious croquettes (cooked by well-known store) and defrosted croquettes were prepared and examined. Fig.10 shows the result of two classes and mean values. It's obvious that the direction of the difference between delicious one and defrosted one parallels to the difference between the cracotte and the apple. Take account of the relationship between the cracotte and the apple, it can be said that the defrosted croquette is wetter crispy food than delicious one.

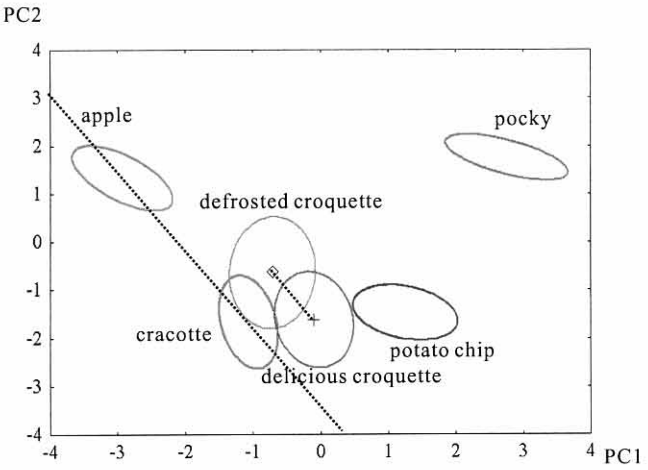


Fig. 10: the comparison of two croquettes

5. Conclusion

In this paper, we proposed a novel acoustic sensor system to evaluate crispness of foods. Since our system mimics bone conduction mechanism in human, minute vibration of foods are detected by teeth, and transmits to left and right sensors through bone. The two sensor outputs give us the localization and spatial spread information as well as our binaural audition. By experiments, we showed such features clearly classified four characteristic crispy foods.

References

[1] Lisa Duizer: "A Review of Acoustic Research for Studying the Sensory Perception of Crisp, Crunchy, and Crackly Texture," *Trends in Food Science & Technology*, Vol. 12, pp. 17–24, 2001.

[2] Z. M. Vickers: "A Psychoacoustical Theory of Crispness," *Journal of Food Science*, Vol. 41, pp. 1158–1164, 1976.

[3] Z. M. Vickers: "Crispness and Crunchiness–A Difference in Pitch?," *Journal of Texture Studies*, Vol. 15, pp. 157–163, 1984.

[4] C. Dacremont: "Spectral Composition of Eating Sounds Generated by Crispy, Crunchy and Crackly Foods," *Journal of Texture Studies*, Vol. 26, pp. 27–43, 1995.

[5] N. De Belie, V. De Smedt, J. De Baerdemaeker: "Principal Component Analysis of Chewing Sounds to Detect Differences in Apple Crispness," *Postharvest Biology and Technology*, Vol. 18, pp. 109–119, 2000.

[6] Z. M. Vickers: "Sensory, Acoustical, and Force-Deformation Measurements of Potato Chip Crispness," *Journal of Food Science*, Vol. 52, No. 1, pp. 138–140, 1987.

[7] Z. M. Vickers: "Instrumental Measures of Crispness and Their Correlation with Sensory Assessment," *Journal of Texture Studies*, Vol. 19, pp. 1–14, 1988.

[8] F. Rohde, M. D. Normand and M. Peleg: "Characterization of the Power Spectrum of Force-Deformation Relationship of Crunchy Foods," *Journal of Texture Studies*, Vol. 24, pp. 45–62, 1993.

[9] S. Ando, H. Shinoda, K. Ogawa, and S. Mitsuyama: "A Three-Dimensional Sound Localization Sensor System Based on the Spatio-Temporal Gradient Method," Trans. Soc. Instrument and Control Engineers (SICE), Vol. 29, No. 5, pp. 520–528, 1993. (in Japanese)