Modeling Sound Generation and Acoustic Characteristics of *KoShik*, a Talking Elephant

Doo-Heon Kyon, Myung-Sook Kim and Myung-Jin Bae

Electronics Engineering Department, Soongsil University
kdhforce@gmail.com

Abstract

This thesis investigates how KoShik, otherwise known as the talking elephant, can have a human-like voice articulation. The mechanism of articulating a human sound by an elephant can be described as a unique excitation process in which the elephant first places its nose in its mouth, blows air into the mouth, produces the basic frequency by pressing the nose with its mouth and teeth, and then articulates the sounds through the narrow airway between the inner skin and soft palate of its mouth. This vocalization process is similar to that of its breeder with respect to the basic pitch and 1-2 formant frequency. The talking elephant KoShik is different from other so-called talking animals in its uniqueness of making a sound by utilizing the nose, more like playing a musical instrument than just mimicking human sounds, as well as of its activeness to talk.

Keywords: *Talking animal, Elephant, Vocalization, Excitation*

1. Introduction

Until now, it has been said that all the sounds produced by the so-called talking animals turned out just a coincidence or a mimicking of a human voice. The case of *KoShik*, the talking elephant of *Everland* (a large amusement park in Seoul, Korea), is considered different from other animals, in his ability to recognize a specific human voice of his breeder, and to use human-like sound articulation very actively.

In the next chapter, we will describe the characteristics of vocalization in a normal elephant, and in the third chapter, we will compare the characteristics of human vocalization and that of *KoShik*, the talking elephant. In the fourth chapter we will show how well KoShik understands the human utterances and in the fifth chapter the findings of this study will be concluded.

2. Characteristics of Elephants' Sounds

Since elephants inhabit in the mother-based society, they are well known for their strong cooperation and companionship. To make a joint society, the elephants gather in various forms, and communicate through sounds.

When they feel comfortable and peaceful, they communicate at a 10Hz(infrasonic), the maximum distance of which is known to cover up to 7~8km. This kind of sound range is unreachable to the human ears. The noise of elephants that we are used to hear, the huge sound like blowing a trumpet, is produced when they feel excited or surprised, or even when they are under attack.

Elephants do not have the evolved tuning organs, unlike humans, so it is impossible for them to vocalize the human words. But there have been reports proving that they can produce human words through experiences and practices by adjusting their vocalization process. A

recorded evidence of one elephant in Kenya mimicking the noise of a truck has been analyzed and its result showed that the elephant can control its vocalization process through hearing experiences[1].

3. Vocalization Model

3.1. The Human Vocalization Model

The acoustic waves of sounds articulated by a human are derived from the shaking of the air through the narrow point of the vocal cord when the air is expelled from the lungs[2].

Sounds can be grouped into three categories according to the types of excitation.

The first type is the voiced excitation producing voiced sounds which are mostly made from the quasiperiodic pulse of air. The air pushed from the glottis with the controlled expansion of the vocal cord stimulates the vocal cord to shake and vibrate.

The second type is the voiceless excitation or stridulation, producing the voiceless sounds which are mostly made from the force of air as it passes, in high speed, through a certain narrow point of the vocal point to make a perturbation. The force of air makes a spectral static in order to stimulate the vocal cord.

The last type is the mixed excitation producing plosive sounds which are made by completely sealing the mouth and building pressure then sudden releasing the air with high pressure [3].

The basic components of acoustic information for each sound are pitch, formant, and amplitude [4]. Even if any two sounds are produced in the same way, their wave forms and spectrums may be changed through a speaker, that is, voice production [5].

3.2 The Vocalization Model of the Talking Elephant

KoShik, a 16 year-old Asian elephant (male), has been reported to pronounce 8 Korean words in a middle aged man's voice.

The flow of air, expelled from the lung, exhales and inhales through the mouth and the long nose. A normal elephant cannot produce complicated pronunciation or sounds due to the lack of a freely moving tongue and an almost undeveloped soft palate which controls the resonance for human voice.

The reason that *KoShik* can speak like a human in spite of these structural limits, is that *KoShik* has a relatively high intelligence in utilizing its nose and mouth properly.

Figure 1 below shows the different positions of the nose when he is in normal status and when he speaks. Normally the nose is out but it is always inside of the moth when he speaks.

(a) Normal   (b) Speaking

Figure 1. Two Figures of *KoShik*

KoShik can speak only by pushing the air from the nose to the mouth. When his nose enters the mouth, it reaches the soft inner side of the mouth. The changes of the excitation are proceeded using the strength of the wind-like air from the nose, pressing the nose with the teeth, and hence controlling the narrow passage in the mouth. The amplitude of the final word was measured as 60~65dB(A).
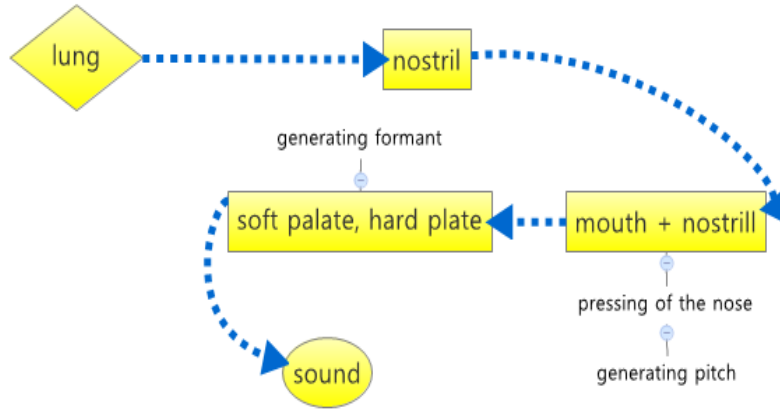
Figure 2. The Vocalization Model of the Talking Elephant

3.3 The Similarities in Pronunciations of the Breeder and *KoShik*

An elephant is usually trained to obey its breeder's commands. When the breeder trained *KoShik*, he mainly used the words for commands shown in the Table 1.

KoShik fully understood the words the breeder used and acted accordingly. Just like the normal elephants, he had the cognitive ability to understand breeder's words and obey him.

The uniqueness of *KoShik* lies in that he not only understood and followed commands, but he also tried to verbally communicate through a special vocalization. The elephant had practiced the words the breeder used, and produced the sounds almost the same.

Table 1. The Words used by the Breeder and the Elephant

The words that breeder use	The vocalization of the elephant
Ha ti	Ah ti
Nu Wah	Nu What
Ah Gik	Ah gi
Jo Hah	Jo hah, Jo ah
An Gah	An Jah
Bal	Bal
-	Yah, Ai
-	Dolla, Di dolla

Figure 3 below is a spectrogram when *KoShik* uses one of the vocalization /Uh Wah/. When comparing the sound patterns of the two with respect to the average frequency of the excitation sound, the breeder shows a 130Hz average and the elephant 132Hz, almost the same. In other words, the elephant's basic tone is very similar to that of the breeder's.

| (a) Breeder | (b) KoShik |
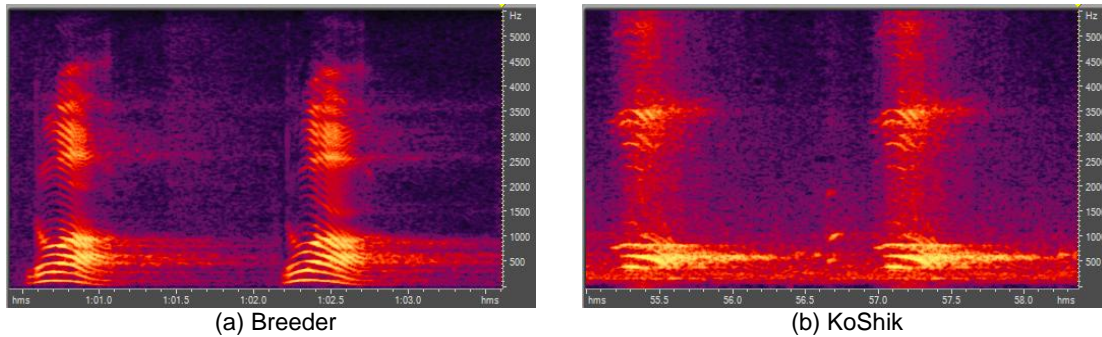
Figure 3. The Spectrograms of the Breeder and KoShik for /Nu wah/

Table 2 below is the formant frequencies of *KoShik* and the breeder. *KoShik*'s first formant shows 550Hz, the breeder 570Hz. The second formant, on average, KoShik measures 2,400Hz and the breeder 2,600Hz, showing a little gap in resonant frequency between the two. The second formant frequency is crucial in recognition, and it is in the range of the first scale (13%). Especially the friction murmur of the third formant in *KoShik* is placed near 3,500Hz, proving that the elephant's vocalization is following the three resonant frequency type which is needed to interact between humans.

It should be also noted that the elephant's vocalization is doing well in pronouncing a diphthong. As shown in the spectrogram for the vocalization of /Nu Wah/, the /u/+/ah/ part shows a special characteristics of spreading in the resonant spectrum. This spreading is clearly shown in both of the breeder's as well as of the elephant's. This means that the elephant has a high intelligence and competence in his own vocalization techniques which is like playing a musical instrument.

Table 2. The Comparison in Formant Frequencies for /Nu wah/

	Breeder (Hz)	KoShik (Hz)
Basic frequency	130	132
First formant	570	550
Second formant	2,650	2,400
Third formant	3,500	3,500

4. The Results of Recognition Test

In order to evaluate the degree of clarity for *KoShik*'s vocalized words, a recognition test was given to find out how well listeners could understand the words produced by *KoShik.*

50 university students in their twenties with no prior knowledge of the given sounds, listened to 7 words which were played three times, and were asked to write down what they heard. The results showed that 76.5% of *KoShik*'s words were successfully delivered. To the question of the identification and the age of the voice, everyone answered a foreigner in their 40s to 50s.

5. Conclusion

This thesis shows how the talking elephant *KoShik* can articulate sounds like a human by explaining the vocalization model. The results show that *KoShik* has overcome his structural

limits of undeveloped tuning organs, and can pronounce words like a human by utilizing his nose and actively interacting with his mouth like a musical instrument.

We also have found out that he not only obeys commands from his breeder but has a deep understanding of the mechanism of his breeder's actual vocalization and thus produces the same sounds. Furthermore, KoShik's pronunciation has been proved human-like as being understood well by normal people.

References

[1] J. H. Poole, P. L. Tyack, A. S. Stoeger-Horwath and S. Watwood, "Elephants are capable of vocal learning", Nature, 434 **(2005)**.

[2] J. L. FlanagenL, "Speech Analysis and Synthesis", Perception, 2nd'edition, Sprniger-Verlag **(1972)**.

[3] M.-J. Bae, "Digital voice coding", DongYoung **(1996)**.

[4] Y.-S. Lee and Y. B. Park, "An acoustic study of feeling information extracting method", The Journal of The Institute of Webcasting, Internet and Telecommunication, 1, 10 **(2010)**.

[5] J.-Y. Ahn, S.-S. Kim, S.-H. Kim, S.-Y. Koh and K.-I. Hur, "A Study on Voice Recognition using Noise Cancel DTW for Noise Environment", The Journal of The Institute of Webcasting, Internet and Telecommunication, 4, 11 **(2011)**.

Authors

Doo-Heon Kyon Ph.D Candidate of Electronic Engineering Department at Soongsil University.

Myung-Sook Kim Ph.D. Professor of English Language and Literature Department at Soongsil University

Myung-Jin Bae Ph.D. Professor of Electronic Information and Communication Engineering Department at Soongsil University