



# Viki's First Words: A Comparative Phonetics Case Study

Axel G. Ekström<sup>1</sup> 

Received: 18 November 2022 / Accepted: 19 January 2023 / Published online: 18 February 2023  
© The Author(s) 2023

It remains an open question whether nonhuman great apes are capable of vocal learning—modification of uttered signals from experience with other individuals. Evidence of such capacities would provide insight into when human imitative abilities—crucial for spoken language—first emerged. However, an important case study with bearing on this topic has been neglected in the literature—that of Viki the chimpanzee (*Pan troglodytes*). Only a few days old, Viki was adopted by Keith and Catherine Hayes of the Yerkes Laboratory of Primate Biology and raised in a human home. Whilst in the Hayeses' care, Viki became an early case in a series of attempts by scientists to teach human language to nonhuman great apes. Ultimately, Viki learned to consistently produce four words: “mama,” “papa,” “cup,” and “up” (Hayes & Hayes, 1951) and was claimed to babble like a human child. This would suggest that chimpanzees are capable of vocal learning. It is surprising, then, that no phonetic analysis of recordings of Viki's speechlike behavior has ever been presented. Providing such an analysis is the purpose of this brief report and case study.

For the study, I curated audio from a segment of the Television movie *The Alphabet Conspiracy* (Warner Bros., 1959). To my knowledge, this film contains the only widely available recording of Viki's speechlike utterances. Across the recording, I identified three pulses of supposed “babble,” three utterances of “papa,” and a single utterance of “cup,” which were all analyzed by way of spectrograms. I was unable to identify the “up” and “mama” utterances purported by the Hayeses (Hayes & Hayes, 1951).

Because of genetic relatedness between chimpanzees and modern humans, I assumed that human approximation of speech-like sounds uttered by a chimpanzee should recruit much of the same articulatory organs and mechanisms. Thus, to compare Viki's speechlike utterances with human phonemes, I, a male speaker aged 26, made recordings imitating Viki's speechlike production. This was done to elucidate potential

---

Handling Editor: Joanna (Jo) M. Setchell

---

✉ Axel G. Ekström  
axeleks@kth.se

<sup>1</sup> Speech, Music & Hearing, KTH Royal Institute of Technology, Lindstedtsvägen 24, 114 28 Stockholm, Sweden

limitations on Viki's articulation. During recordings, I listened to each of Viki's utterances and vocally approximated the sound heard. Each of Viki's utterances was imitated three times and the approximation closest to that of Viki (to the ear) was rendered and analysed. All sounds were rendered as high-pass filter spectrograms (300 Hz) in the *Sopran* software (Svante Granqvist, Karolinska Institute) (tolvan.com) (Table 1).

## Babble

Human infants typically begin exhibiting babbling behavior around the age of 6 months. In developmental sciences, babble is defined as syllabic prespeech behavior and is widely believed to lay the groundwork for speech proper. If indeed chimpanzees were capable of vocal learning in the wild (Crockford *et al.*, 2004), it is conceivable that infant chimpanzees should occasionally babble also (but see Kojima, 2001). The sequence of supposed babble by Viki consists of three pulses in quick succession, which I realized as glottal stops [ʔ] - formed via airstream release following complete occlusion by the glottis (“uh-oh”). However, Viki's so-called babbling does not exhibit the telltale acoustic features of either syllabic human speech (MacNeilage, 1998), nor does it resemble her other speech-like utterances. Given its singular occurrence, and the controversial nature of the claim, Viki's supposed babbling is more parsimoniously interpreted as playful grunting.

## Papa

Viki's /papa/ lacks any apparent vowel-like features, which are readily identifiable by their arrangement as broad bands of concentrated energy. Instead, utterances are arranged as sequences of two clicks or bursts (consistent with plosive consonants). However, the acoustic profiles of segments were less consistent with bilabial plosive [p] (“pot”), and more with tenuis (voiceless, unaspirated, unglottalized obstruents) bilabial clicks [⦿] (rare in natural languages, occurring only in the Tuu and Kx'a language families of southern Africa).

Crucially, across all iterations of /papa/, the spectrographic profile of the utterance is well maintained, indeed suggesting a learned motor sequence (i.e., a word). Importantly, consistent execution of the corresponding speech gestural score indicates that Viki had indeed acquired a novel voiceless call. Because vocal learning typically refers to voiced sounds, such learned gesture may be tentatively described as *nonphonatory* vocal learning, reflecting the lack of laryngeal involvement.

## Cup

As with, /papa/, there is no apparent vowel-like articulation following /c/ in Viki's /cup/. Furthermore, the spectrographic profile of the /c/ in cup was not consistent with voiceless velar plosive [k] but more so with my production of voiceless velar fricative [x] (German: “Bach”). In humans, velar consonants are executed via the dorsal tongue against the velum—but [x] is executed posteriorly in comparison. Viki's potential realization of [x] (a fricative) as opposed to [k] (a plosive)

**Table 1** Transcriptions and spectrograms of Wiki's speechlike productions, and mode of articulation (in humans, application to Wiki is tentative). In all spectrograms, time (in seconds) is represented on X axis; frequency in Hertz (Hz) on Y axis

Utterance	Phonetic transcription	Features of phoneme	Recruits*
Babble**	/ʔ/	[ʔ]: Complete occlusion and subsequent release of airstream.	Glottis.
<i>Wiki</i>			
<i>Human</i>			
"Papa"	/OO/	[O]: Ingressive trapping of air pocket between closures (lips), and subsequent release.	Upper and lower lips.
<i>Wiki</i>			
<i>Human</i>			
"Cup"	/xp/	[x]: Constriction of airflow causing turbulence at place of articulation. [p]: Complete (bilabial) occlusion of airflow, followed by release.	Dorsum (pharyngeal or posterior tongue) and velum (soft palate). Upper and lower lips.
<i>Wiki</i>			
<i>Human</i>			

may be suggestive of limitations of the chimpanzee tongue, which is longer and flat in comparison with humans, with fewer articulatory degrees of freedom inside the oral cavity accordingly (Takemoto, 2008). However, the interpretation of Viki's /c/ as [x] should be considered tentative, as intra-oral articulatory gestures actively employed by chimpanzees in call production remain poorly understood.

The /p/ in “cup” appears similarly to the /p/ in Viki's “papa”—but with an evidently weaker release (diminished in intensity). This is consistent with the interpretation of the /p/ as [p]. I also realized it as such.

The cohesive gestural score of [xp], and the temporal closeness between phonemes [x] and [p] is again suggestive of a learned sequence (nonphonatory vocal learning), although without repeated iterations, this cannot be readily determined. Both [x] and [p] are voiceless.

## Summary

I have provided phonetic analyses of Vicki's speechlike utterances and argued that they align acoustically with human production of voiceless bilabial plosive [p], voiceless velar fricative [x] or voiceless pharyngeal nonsibilant fricative [ħ], voiceless glottal stop [ʔ], and voiceless bilabial click [ǀ]. This indicates that (1) articulatory capacities of chimpanzees may ultimately be defined by species-typical tongue dimensions and degrees of freedom (Takemoto, 2008), and (2) at least in rare occurrences, chimpanzees may be capable of *nonphonatory* or voiceless vocal learning. While Viki was seemingly incapable of simultaneous recruitment of consonantal “frames” and vowel-like “content” (MacNeilage, 1998), she successfully produced a small sample of humanlike consonantal speech sounds, including labial and, possibly, velar articulations.

**Acknowledgements** I am grateful to Andrey Anikin and to an additional anonymous reviewer for insightful commentary on the text and its presentation. Additional thanks to Steven Moran and Adriano Lameira for continued discussion on ape articulation.

**Funding** Open access funding provided by Royal Institute of Technology.

**Data Availability** All .wav files analyzed in the study are publicly available ([github.com/evofant/Viki\\_speech](https://github.com/evofant/Viki_speech)).

**Ethical Note** The study in its entirety was carried out in line with guidelines issued by the Swedish Ethical Review Authority ([Etikprövningsmyndigheten.se](http://etikprovningsmyndigheten.se)) for research on animal subjects.

## Declarations

**Conflict of Interest** The author declares no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Crockford, C., Herbinger, I., Vigilant, L., & Boesch, C. (2004). Wild chimpanzees produce group-specific calls: a case for vocal learning? *Ethology*, *110*(3), 221–243.
- Hayes, K. J., & Hayes, C. (1951). The intellectual development of a home-raised chimpanzee. *Proceedings of the American Philosophical Society*, *95*(2), 105–109.
- Kojima, S. (2001). Early vocal development in a chimpanzee infant. In T. Matsuzawa (ed.), *Primate origins of human cognition and behavior* (pp. 190–196). Springer, Tokyo.
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, *21*(4), 499–511.
- Takemoto, H. (2008). Morphological analyses and 3D modeling of the tongue musculature of the chimpanzee (*Pan troglodytes*). *American Journal of Primatology: Official Journal of the American Society of Primatologists*, *70*(10), 966–975.