

# The Zurich Corpus of Vowel and Voice Quality

Version 2

Dieter Maurer, Christian d'Heureuse, Heidi Suter,  
Volker Dellwo, Daniel Friedrichs, Thayabaran Kathiresan

## Vowel Imitations by Birds

(Last update 2025-02-28)

### Part I

Excerpt of the book 'Acoustics of the Vowel – Preliminaries' (Maurer, 2018, Chapters 10.A and M10.A)

(Preliminaries, Chapter 10.A, pp. 68–69:)

“Sounds of animals imitating utterances of humans are also of primary importance in the discussion of vowel sounds, related spectral characteristics, formant patterns, perceived speaker groups and vocal-tract sizes.

“Fundamental in this respect is the question of how birds are able to imitate human sounds despite lacking the means of phonation and articulation—in particular, a corresponding vocal-tract resonance.

“According to our own preliminary examination of vowel imitation by common hill myna birds who excel at such mimicry (results unpublished, although some clear examples are given in the Materials section), we conclude the following: if these birds imitate words, and if individual imitated vowel sounds are isolated as sound fragments in a way that they possess a quasi-static character in terms of quasi-static spectral characteristics (above all, that transitions are excluded), then vowel perception and a distinction of such sounds by humans is possible. For part of these sound fragments, complete F1–F2–F3 formant patterns comparable to patterns given for human sounds can be interpreted. For the remaining fragments, only a partial correspondence in formant patterns can be observed. (However, this statement must be relativised: strictly speaking, any calculation of vowel-related formant patterns of bird sounds is methodically unsubstantiated; see below.)

“The fact that birds are able to imitate human vowel sounds with vowel-specific spectral characteristics and formant patterns comparable to those of humans contradicts, in its turn, a strict correspondence between the spectral characteristics of the produced sound and vocal-tract resonance. The same holds true for a strict correspondence between spectral characteristics of the produced sound and vocal-tract size. Consequently, any critical investigation and discussion of vowels must focus on the possibility that the same sound characteristics can be produced under substantially different physical and physiological conditions.

“Besides, if birds are able to mimic human utterances, they must be able to perceptually differentiate different vocal sounds. However, their perception cannot rely on any sensomotoric and conceptual experience of vowel production comparable to the experience of humans. Thus, it can be speculated that their perception relies on a more “abstract” acoustic “form” of the vowel sound. (Such speculation would meet the claim that a phenomenological approach to the physical representation of vowels is needed; see Part V.)”

(Preliminaries, Chapter M10.A, p. 238:)

“The following series show examples of vowel sounds of common hill mynah birds (*Gracula religiosa*) imitating vocal expressions and words of humans. The examples are selected on the basis of extensive recordings of 21 birds, most of them living in Indonesia. (However, they imitated words of different languages.) The spectra presented relate to vowel nuclei extracted from the expressions or words. Both the entire imitated expressions or words as well as the extracted sound fragments are perceptually recognisable.

“In each of the series, the sound spectra are given in the order of the birds and of F0. (Note that in several cases, different sound spectra for the same vowel are shown for a bird, in order to document variations in F0 and the sound spectra.) – [Please note:] Acoustic analysis corresponds to the analysis as described in the Note on the Method section. LPC filter curves relate to a parameter setting of the LPC analysis according to the PRAAT standard for women. However, as mentioned in the text, the LPC analysis is not methodically substantiated.”

“**Figure 11:** Examples of sounds of imitated /i/ in word context produced by five birds, with F0 ranging from c. 110–380 Hz; perceptual vowel quality is /i/, including intermediate qualities /i–j/, /i–y/ and /i–e/.” (Preliminaries, pp. 238 and 240)

All sounds:

<https://zhcorpus.org/v2/db/query?obj=218968+217524+218126+218037+218789+218859+218857+219334+219346>

Two clear examples with the entire words (word+vowel+word+vowel):

<https://zhcorpus.org/v2/db/query?obj=218829+218859+219138+219346>

Note:

218859 = HP 1 kHz = i, LP 1 kHz = o

219346 = HP 1 kHz = i, LP 1 kHz = u

“**Figure 12:** Examples of sounds of imitated /e/ in word context produced by five birds, with F0 ranging from c. 160–330 Hz; perceptual vowel quality is /e/, including intermediate qualities /e–i/ and /e–ø/.” (Preliminaries, pp. 238 and 241)

All sounds:

<https://zhcorpus.org/v2/db/query?obj=218072+218144+218124+218808+218777+218778+218768+218765+218854+218847+218873>

Three clear examples with the entire words (word+vowel+word+vowel+...):

<https://zhcorpus.org/v2/db/query?obj=217925+218072+218751+218808+218706+218778>

Note:

218072 = HP 1 kHz = e (energy < 1 kHz with no substantial impact on vowel quality); HP 2 kHz = towards /i/ (open–close vowel quality shift); LP 1 kHz = //openo//

218808 = HP 1 kHz = not clear; HP 2 kHz = /i/ (open–close vowel quality shift); LP 1 kHz = /o/

218778 = HP 1 kHz = not clear; HP 2 kHz = /i/ (open–close vowel quality shift); LP 1 kHz = /o/

**“Figure 13:** Examples of sounds of imitated /a/ in word context produced by twelve birds, with F0 ranging from c. 110–490 Hz; perceptual vowel quality is /a–a/, including intermediate quality /a–ɔ/.” (Preliminaries, pp. 238 and 242–244)

All sounds:

<https://zhcorpus.org/v2/db/query?obj=218974+218976+218982+217494+217846+218263+218188+218102+218193+218121+218241+218234+218467+218680+218614+218584+218566+218654+218795+218772+218773+218782+219031+219109+219349+219344+219345+219310+219309+219304+219300+219296+219290>

Note:

Some sounds are ‘rough’ (creaky), others are clearly voiced (compare e.g. 218992 and 218680).

In some sound spectra, the harmonic series is ‘incomplete’ or the amplitudes of the harmonics vary strongly.

In many sounds, the amplitude of H1 is low or H1 is even completely absent.

Three types of spectral characteristics in the presumed vowel-related frequency range < 1.5 kHz are observed: two spectral peaks, only one peak and a flat spectral part.

Five clear examples with the entire words (word+vowel+word+vowel+...):

<https://zhcorpus.org/v2/db/query?obj=217401+217494+217998+218263+218344+218566+218380+218614+219202+219300>

**“Figure 14:** Examples of sounds of imitated /o/ in word context produced by eleven birds, with F0 ranging from c. 80–410 Hz; perceptual vowel quality is /o /, including intermediate quality /o–ɔ/.” (Preliminaries, pp. 238 and 245–246)

All sounds:

<https://zhcorpus.org/v2/db/query?obj=217461+217465+217477+217497+218949+217499+217829+218212+218579+218611+218640+218637+218631+218781+218878+218875+219034+219060+219051+219108+219335+219331+219320+219329>

Note:

Some sounds are ‘rough’ (creaky-like), some are clearly voiced-like (compare e.g. 219108 and 219051).

For many sounds, the amplitude of H1 is low or H1 is even absent.

Two types spectral characteristics in the presumed vowel-related frequency range < 1.5 kHz are observed: two spectral peaks or only one peak.

Six selected vowel sounds with ascending fo level (and corresponding pitch level) from 81 Hz to 411 Hz:

<https://zhcorpus.org/v2/db/query?obj=217461+219034+217465+219051+218631+217477>

Note:

- For the sound 217461 (fo = 81 Hz), the estimated value for F1-F2 is 451-1054 Hz, but for the sound 219051 (fo = 280 Hz), the estimated value for F1-F2 is 639-1229 Hz (see the calculated values for parameter setting P6, and compare the spectrograms). – At the same time, for the sound 217461, vowel quality is somewhat maintained in the /o–ɔ/ boundary in the Klatt resynthesis (based on the above F(i) values related to the LPC parameter setting P6, where the bandwidth of F1 is set to 50 Hz) applying original fo = 81 Hz, but the vowel quality shifts towards /u/ when fo is set to 411 Hz (open–close vowel quality shift with ascending fo; AE); for the sound 219051, the vowel quality is also somewhat maintained in the /o–ɔ/ boundary in the Klatt resynthesis (based on the above F(i) values related to the LPC parameter setting P6) applying original fo = 280 Hz, but the vowel quality shifts to /a/ if fo is set to 81 Hz (close–open vowel quality shift with descending fo; AE).

- For the sound 217461 ( $f_0 = 81$  Hz), the vowel spectrum shows a relative energy minimum in the frequency region of 700–800 Hz, whereas for the sound 218631 it shows an energy maximum; this corresponds to the observation of inversions of vowel-related relative spectral energy maxima and minima (see Preliminaries, Chapter M8.3, and Indices, Chapter 7.2 and M7.2).
- In these terms, the relation of the lower vowel spectrum to  $f_0$  is mirrored in the vocal mimicry.
- For corresponding sound comparisons, see also the sounds of /u/ documented below.

Six clear examples with the entire words (word+vowel+word+vowel+...):

<https://zhcorpus.org/v2/db/query?obj=217371+217461+219003+219034+217375+217465+219020+219051+218400+218631+217384+217477>

Note:

- For sound 218631, the vowel quality is maintained in harmonic synthesis (HarmSyn) with only H2 and H3 enabled.

**“Figure 15:** Examples of sounds of imitated /u/ in word context produced by seven birds, with  $F_0$  ranging from c. 110–660 Hz; perceptual vowel quality is /u/, including intermediate quality /u–o/.” (Preliminaries, pp. 238 and 247–248)

All sounds:

<https://zhcorpus.org/v2/db/query?obj=217452+217443+218969+218973+217478+217752+217755+217779+218169+218531+218590+218818+218866+218851+219293>

Six selected vowel sounds with ascending  $f_0$  level (and corresponding pitch level) from 105 Hz to 651 Hz:

<https://zhcorpus.org/v2/db/query?obj=217752+219293+217452+218851+218590+217478>

Note:

- Similar to human vowel sounds, the spectra show pronounced upward shifts of prominent energy < 1 kHz in relation to the ascending  $f_0$ .

Two clear examples with the entire words (word+vowel+word+vowel):

<https://zhcorpus.org/v2/db/query?obj=217365+217452+218821+218851>

### Additional note

“Note that many of the sound spectra of these birds are similar to the vowel spectra of humans presented in the previous sections. However, for some examples of imitations of front vowels, the lower part of the spectral configuration < 1 kHz is unexpected’.” (Preliminaries, p. 239)