Reconstructing Basay Vowel Acoustics from Asai Erin's Archival Recordings

— Praat Formant Measurement and eSpeak-NG Synthesis for an Extinct Language —

Author: Tsai Yung-kuei (蔡永桂)

Date: April 2026

Type: Original research (acoustic phonology / language documentation)

 Citation ID: basay.tw/research/2026-04-basay-acoustic/

Abstract

This paper presents the first formant-measurement-based acoustic reconstruction of the Basay (Basai) vowel system, an extinct Formosan Austronesian language of northeastern Taiwan. Formant frequencies F1, F2, and F3 for six vowels were measured directly from archival audio recorded by Japanese linguist Asai Erin (ca. 1936) from a Trobiawan (哆囉美遠) dialect speaker, using the Praat acoustic analysis software. The recordings, originally held at the Museum of Anthropology, National Taiwan University, were obtained via the Tokyo University of Foreign Studies Language Archive. The measured formant values were implemented as a working speech

synthesizer by reverse-engineering the SPECTSQ2 binary format of eSpeak-NG 1.52 and developing Python tooling. Comparison with previously used transcription-based phoneme approximations reveals substantial differences — notably /o/ (F1: +80.5%) and /a/ (F2: +30.5%) — demonstrating that transcription-based resynthesis cannot reliably represent the acoustic properties documented in the original recordings. The synthesized vowels were validated by spectrogram analysis, and the methodology provides a replicable framework for evidence-based acoustic reconstruction of extinct languages.

Keywords: Basay, Formosan, acoustic phonology, formant measurement, Praat, eSpeak-NG, archival recordings, extinct language documentation

Cite this article

APA:

```
Tsai, Y.-K. (2026). Reconstructing Basay vowel acoustics from Asai Erin's archival recordings: Praat formant measurement and eSpeak-NG synthesis for an extinct language. basay.tw. https://basay.tw/research/2026-04-basay-acoustic/
```

BibTeX:

```
@misc{tsai2026basay,
  author = {Tsai, Yung-kuei and 蔡永桂},
  title  = {Reconstructing {Basay} Vowel Acoustics from {Asai Erin}'s Archival Recordings},
  year   = {2026},
  month  = {4},
  url    = {https://basay.tw/research/2026-04-basay-acoustic/},
  note   = {Praat formant measurement and eSpeak-NG synthesis
```

```
for an extinct language}
}
```

---

1. Introduction

Basay (also spelled Basai) is an extinct Austronesian language historically spoken along the northeastern coast of Taiwan, in the area of present-day Taipei, Keelung, and the Yilan Plain. Linguistically, Basay belongs to the Formosan branch of the Austronesian family and is most closely related to Kavalan, the two languages constituting the only members of the Northeast Formosan subgroup. The last fluent native speakers of Basay were documented in the 1930s by Japanese linguist Asai Erin (Asai 1937), whose field notes and phonograph recordings constitute the primary — and irreplaceable — documentation of the language.

Since Asai's documentation, subsequent research — including the foundational studies by Li Paul Jen-kuei (Li 1993, 1999, 2001) and the comparative reconstruction by Tseng Li-yang (Tseng 2022) — has focused primarily on phonological transcription, grammatical analysis, and lexical reconstruction. These contributions are invaluable; however, none have made full use of the most fundamental resource in Asai's archival recordings: the acoustic signal itself. To our knowledge, no prior work has extracted formant measurements directly from these recordings, leaving the acoustic properties of Basay vowels unquantified.

This study fills that gap. We use Praat acoustic analysis software to measure F1, F2, and F3 formant frequencies directly from Asai's

recordings of a Trobiawan (哆囉美遠) dialect speaker. These empirically grounded values are then implemented as a working text-to-speech synthesizer by reverse-engineering the SPECTSQ2 binary format of the eSpeak-NG 1.52 engine and developing Python tooling. The result is the first phonetically grounded acoustic model of Basay vowels.

This study also demonstrates the inadequacy of the previous approach — selecting eSpeak phoneme files by transcription symbol similarity — by comparing those files' formant values against our measured targets. The differences are acoustically substantial, indicating that transcription-based approximation is not a substitute for measurement-based reconstruction.

2. Background and Prior Work

2.1 Asai Erin's Recordings

The primary acoustic resource for this study is a set of phonograph recordings made by Asai Erin during fieldwork in Taiwan (ca. 1936). The recordings document speakers of the Trobiawan (哆囉美遠) dialect of Basay from the Yilan region. The original recordings are housed at the Museum of Anthropology, National Taiwan University; copies were obtained for this study via the Tokyo University of Foreign Studies Language Archive. The recordings include narrative texts, among them "Story of Ngazi (3)" and "Story of Mutravai (1)".

The Trobiawan dialect represents one of the two documented dialect areas of Basay. Asai (1937) notes that his primary Trobiawan informant, Mutravai, also spoke Kavalan, and the possibility of some phonetic transfer cannot be excluded. This language contact

situation is acknowledged as a limitation on the representativeness of the acoustic data, but it does not diminish the recordings' importance as the only existing acoustic record of Basay speech.

2.2 Prior Language Documentation

Asai Erin's 1937 field notes provide the primary transcription record of Basay phonology. Li Paul Jen-kuei (1999) provides the most comprehensive phonological analysis to date, identifying approximately 15 consonant phonemes and a vowel system of 4–6 vowels. Tseng Li-yang (2022), in a comparison of Basay and Kavalan, proposes a reconstruction of Proto-Northeast Formosan with four vowels (*i, *u, *ə, *a) and four diphthongs.

None of these studies attempted acoustic formant measurement of the vowels. This study provides such measurements directly from the archival recordings, rather than relying on inference from transcription.

2.3 eSpeak-NG and Formant Synthesis

eSpeak-NG is an open-source text-to-speech engine using a Klatt-based formant synthesizer. Vowel quality is specified via binary files in the SPECTSQ2 format, which encodes spectral frames containing formant frequency, bandwidth, and amplitude parameters. The synthesis engine reads formant values directly from these binary files (`peaks[1]`, `peaks[2]`, `peaks[3]` for F1, F2, F3 respectively), permitting formant specification at Hz-level precision. For this study, Python tooling was developed to generate SPECTSQ2-format files directly from measured formant values, enabling a reproducible, platform-independent synthesis pipeline.

3. Methods

3.1 Acoustic Measurement

Formant frequencies were measured using Praat (Boersma & Weenink) from Asai's archival recordings of the Trobiawan dialect, specifically the narrative texts "Story of Ngazi (3)" and "Story of Mutravai (1)". Steady-state portions of vowel tokens were identified in spectrogram view, and F1, F2, F3 values were extracted at the midpoint of each token using Praat's formant tracker. The resulting measurements represent the acoustic realization of Basay vowels as produced by the Trobiawan dialect speaker recorded by Asai around 1936. A third recording, "Story of Mutravai (4)", was used separately for methodological development of an automated vowel classification procedure; it did not contribute to the formant values implemented in the synthesizer.

To assess the reliability of formant measurements and examine the potential influence of recording noise, an additional fourth recording was analyzed: "Story of Saturai (1)", a sung recording from which 9,991 analysis frames were extracted. Across the three sources — the two prose narratives used for implementation and the sung recording — coefficients of variation (CV) for F2 were below 10% for all six vowels (range: 1.9–6.0%), and below 5% for F1 for four vowels (/e/, /a/, /o/, /ə/). The two vowels with higher F1 CVs (/i/: 16.3%; /u/: 12.2%) had fewer tokens in the prose recordings (N=40 and N=174 respectively), suggesting statistical sampling effects rather than acoustic measurement error. These results indicate that formant measurements are robust across recording sessions and speech

styles, and that 1930s recording noise does not materially affect the measurements.

Bandwidth estimates (B1, B2, B3) were set according to standard modal phonation norms consistent with the measured formant configuration, as bandwidth information cannot be reliably extracted from the archival recording quality. Measured formant values and bandwidth estimates are given in Table 1.

Table 1. Basay vowel formant frequencies (Trobiawan dialect). F1–F3 measured using Praat from Asai's archival recordings (ca. 1936). Bandwidths (B1–B3) are standard modal phonation estimates.

Vowel (IPA)	F1 (Hz)	F2 (Hz)	F3 (Hz)	B1 (Hz)	B2 (Hz)	B3 (Hz)	Source
/i/	269	1696	2128	60	80	120	Measured
/e/	639	1701	2245	70	90	120	Measured
/ə/ (central)	858	1703	2180	100	100	120	Measured
/a/	1064	1671	2185	120	110	130	Measured
/o/	805	1191	1949	80	100	130	Measured
/u/	381	1219	1910	60	90	130	Measured

3.2 Reverse Engineering the SPECTSQ2 Format

The SPECTSQ2 binary format was reverse-engineered by analyzing the eSpeak-NG source code (`spect.cpp`, `spect.h`, `synthesize.h`)

and cross-referencing hex dumps of existing vowel files. Key findings include:

- The file header uses the magic signature `SPECTSQ2` (0x53504543 + 0x54535132), followed by a little-endian length-prefixed name string.
- Floating-point values (time, F0, length, dx) are stored in **80-bit extended precision** (10 bytes each), not standard IEEE 754 64-bit double.
- F1, F2, F3 are encoded in `peaks[1]`, `peaks[2]`, `peaks[3]` respectively; `peaks[0]` encodes the F0 region.
- `KLATT_Kopen` (`klatt_param[]` index 5) must be set to 50 to ensure clean voiced modal synthesis without frication noise.

3.3 Binary File Generation

SPECTSQ2 files for the six Basay vowels were generated by a Python script (`gen_basay_vowels_v8.py`), patching existing eSpeak-NG vowel files as templates. The patching procedure: (1) replaces `peaks[1–3].pkfreq` and `formants[1–3].freq` with the measured F1–F3 values; (2) zeros `peaks[0].pkheight` and `peaks[4–8].pkheight` to suppress non-formant spectral components; (3) zeros the harmonic spectrum array (`spect[]`) to eliminate spectral coloring inherited from the template vowel. The `KLATT_Kopen` value and other synthesis parameters are preserved from the template to ensure clean voiced output.

4. Results

4.1 Comparison with Prior Transcription Approximations

Prior to this acoustic measurement, Basay vowels in eSpeak had been approximated using existing phoneme files selected by transcription symbol similarity (e.g., /i/ → `i_fnt`, /o/ → `o_mid`). Table 2 compares the formant values of these approximation files against the values measured from Asai's recordings.

Table 2. Comparison of prior transcription approximations vs. measured values from Asai's archival recordings. *`i_fnt` uses the older SPECTSEQ format, which lacks key-frame height data.

Vowel	Parameter	Prior approx. (Hz)	Measured (Hz)	Difference (Hz)	Difference (%)
/i/	F1	~0 *	269	—	—
	F2	~0 *	1696	—	—
	F3	~0 *	2128	—	—
/e/	F1	584	639	+55	+9.4%
	F2	1820	1701	−119	−6.5%
	F3	2560	2245	−315	−12.3%
/ə/	F1	581	858	+277	+47.7%
	F2	1653	1703	+50	+3.0%

Vowel	Parameter	Prior approx. (Hz)	Measured (Hz)	Difference (Hz)	Difference (%)
	F3	2477	2180	−297	−12.0%
/a/	F1	872	1064	+192	+22.0%
	F2	1280	1671	+391	+30.5%
	F3	2660	2185	−475	−17.9%
/o/	F1	446	805	+359	+80.5%
	F2	883	1191	+308	+34.9%
	F3	2485	1949	−536	−21.6%
/u/	F1	371	381	+10	+2.7%
	F2	1276	1219	−57	−4.5%
	F3	2308	1910	−398	−17.2%

The largest discrepancy is for /o/ (F1: +80.5%, F2: +34.9%), followed by /ə/ (F1: +47.7%) and /a/ (F2: +30.5%). The 359 Hz F1 gap for /o/ is acoustically significant: the approximation file placed this vowel in the mid-vowel range (446 Hz), while the measured value (805 Hz) identifies it as a higher mid-back vowel. F3 values for back vowels (/o/, /u/) were systematically overestimated by 400–536 Hz in prior approximations, a consequence of using front-vowel templates.

The case of /i/ is most extreme: the previously used approximation file (`i_fnt`) employs the older SPECTSEQ format, which lacks key-

frame height data, leaving this vowel acoustically undefined prior to this reconstruction. The current implementation of /i/ in the Basay synthesizer is based entirely on measured values from Asai's recordings (F1: 269 Hz, F2: 1696 Hz, F3: 2128 Hz), with no valid prior approximation to compare against. This further illustrates the fundamental inadequacy of the transcription-approximation approach: for /i/ — typologically the most canonical and cross-linguistically most stable high front vowel — the previous method was unable to produce any usable acoustic output at all.

4.2 Spectrogram Validation

Synthesized vowels were validated by spectrogram analysis using SoX. Spectrograms confirm that F1 ordering follows the expected typological hierarchy: F1(/a/) > F1(/o/) > F1(/e/) > F1(/u/) > F1(/i/), consistent with the universal vowel height dimension. The front-back contrast is also correctly represented: /i/, /e/, /ə/ have F2 values clustered near 1700 Hz, while /o/ and /u/ cluster near 1200 Hz, consistent with the measured values and the expected back-vowel character of /o/ and /u/.

All synthesized vowels show residual broadband noise above 4 kHz; analysis indicates this is a property of the Klatt synthesis engine's voicing model rather than a problem with formant specification, and does not affect the acoustic validity of the F1–F3 implementation.

5. Discussion

The core contribution of this study is methodological: we demonstrate that acoustic measurement from archival recordings — even recordings of limited quality — yields formant values that differ

substantially and systematically from transcription-based approximations. The discrepancies documented in Table 2 are not random variation; they reflect specific structural failures of the approximation approach: using `i_fnt` for /i/ produced a file with no usable formant data; using `o_mid` for /o/ underestimated F1 by 359 Hz, misrepresenting vowel height; using front-vowel templates for /o/ and /u/ systematically overestimated F3 by 400–536 Hz.

None of these failures would have been detectable without returning to the original recordings. This study demonstrates that Asai's archival recordings — though not collected for acoustic phonological purposes — contain signal quality sufficient for Praat-based formant extraction. This finding has implications beyond Basay: it suggests that archival recordings of other extinct Formosan languages, where they exist, may be similarly amenable to acoustic measurement.

Several limitations apply to the measurements in this study. First, the recorded speaker (Trobiawan dialect) also spoke Kavalan, and phonetic transfer cannot be entirely excluded. Second, the exact recording date (ca. 1936) remains uncertain. Third, bandwidth values were not measured from the recordings but estimated from standard modal phonation norms. Future work should attempt direct bandwidth measurement, and where suitable recordings can be identified, compare Trobiawan formant values with those of the Xinzhe (新社) dialect.

To assess measurement reliability, three independent recordings were analyzed: the two prose narratives used for formant implementation (Story of Ngazi (3) and Story of Mutravai (1)) and an additional song recording (Story of Saturai (1), N=9,991 frames). Coefficients of variation across the three sources were below 10% for F2 for all six vowels (range: 1.9–6.0%), and below 5% for F1 for four

vowels (/e/, /a/, /o/, /ə/). The two vowels with higher F1 variation (/i/: 16.3%; /u/: 12.2%) had fewer tokens in the prose recordings (N=40 and N=174 respectively), suggesting sampling effects rather than recording noise. These results indicate that formant measurements are robust across recording contexts and speech styles, and that recording noise does not materially affect the measurements.

One parameter not yet implemented is lexical stress. Stress position in Basay cannot be determined from the available recordings with sufficient confidence: the archival audio does not allow reliable identification of stressed syllables by acoustic correlates such as F0 rise, duration, or amplitude. Preliminary observation suggests a pattern of final stress in disyllables and penultimate stress in trisyllables, consistent with common Austronesian stress templates, but this remains a hypothesis rather than an empirically confirmed rule. The current implementation uses eSpeak-NG's default stress assignment. Source code inspection confirms that the `stressrule` parameter is not implemented in eSpeak-NG 1.52's language configuration, so stress assignment cannot be controlled through the lang file alone. Once more reliable acoustic evidence becomes available — ideally through systematic measurement of duration and F0 across multiple tokens of the same words in the archival recordings — this parameter should be revised.

The eSpeak-NG implementation described in this paper is fully reproducible. The Python generation tooling, SPECTSQ2 format documentation, and formant target values are available to the research community, and the synthesizer can be updated as new acoustic evidence emerges.

A particularly noteworthy finding concerns the near-merger of /ə/ with /a/ and /o/ in both measured formant values and synthesized

output. The F2 values of all six vowels are systematically compressed into a narrow posterior range (1191–1703 Hz), with /a/, /o/, and /ə/ especially close. This compression is consistent with at least three mutually reinforcing factors. First, phonatory drift in elderly speakers is well documented, including retraction of front and central vowels, pharyngeal expansion, and reduced labial constriction — all of which reduce the F2 range of the vowel space. Second, the recorded speaker also spoke Kavalan, whose reconstructed vowel inventory (*i, *u, *ə, *a) does not include a distinct /o/ or /e/, potentially weakening the phonemic contrast between /ə/ and adjacent vowels in this speaker. Third, Tseng's (2022) reconstruction of Proto-Basay includes only four vowels (*i, *u, *ə, *a), suggesting that Basay /e/ and /o/ may represent later expansion or contact borrowing with correspondingly weaker phonemic entrenchment. The perceptual near-merger of /ə/ with /a/ and /o/ in the synthesized vowels may therefore faithfully represent a genuine weakening of phonemic contrasts in the original recordings, rather than a limitation of the synthesis method. This observation is itself a phonological finding: the archival recordings preserve acoustic evidence of vowel space compression, consistent with what has been documented in the phonetics of terminal speakers of other extinct and endangered languages.

6. Conclusion

This study presents the first acoustic reconstruction of the Basay vowel system based on direct formant measurement from Asai Erin's archival recordings (ca. 1936, Trobiawan dialect, obtained via Tokyo University of Foreign Studies). Formant frequencies F1, F2, and F3 for six vowels were extracted using Praat and implemented in eSpeak-

NG 1.52 via a reverse-engineered binary format and Python synthesis tooling. Comparison with prior transcription approximations reveals substantial discrepancies that would not have been detectable without returning to the original recordings. The resulting synthesizer constitutes the first phonetically grounded acoustic model of Basay, and the methodology provides a replicable template for evidence-based acoustic reconstruction of extinct languages with surviving archival recordings.

References

- Asai, Erin (淺井惠倫). 1937. Basay field notes. [Recordings ca. 1936, Trobiawan dialect.] Originally held at the Museum of Anthropology, National Taiwan University; copy at the Tokyo University of Foreign Studies Language Archive.
- Boersma, Paul & David Weenink. Praat: doing phonetics by computer. www.praat.org
- Li, Paul Jen-kuei (李壬癸). 1993. New data on three extinct Formosan languages. *Bulletin of the Institute of History and Philology, Academia Sinica* 63(2): 301–323.
- Li, Paul Jen-kuei (李壬癸). 1999. Some problems in Basay. In E. Zeitoun & P.J.K. Li (eds.), *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, 635–664. Taipei: Academia Sinica.
- Li, Paul Jen-kuei (李壬癸). 2001. The linguistic status of Basay. *Language and Linguistics* 2(2): 155–171.
- Tseng, Li-yang (曾立洋). 2022. Reconstruction of Northeast Formosan. Master's thesis.
- Dunn, Reece H. et al. eSpeak-NG. github.com/espeak-ng/espeak-ng

← Back to Research