

# 從淺井惠倫檔案錄音重建巴賽語母音音響特性： 以 Praat 共振峰測量與 eSpeak-NG 合成重現消滅語言 語音

蔡永桂

2026 年 4 月

## 摘要

本文提出巴賽語 (Basay/Basai) 母音系統的首次以共振峰測量為基礎之音響重建。巴賽語為臺灣東北部已消滅的南島語系福爾摩沙語族語言。六個母音的共振峰頻率 (F1、F2、F3)，係直接從日本語言學家淺井惠倫 (約 1936 年) 針對哆囉美遠 (Trobiawan) 方言說話者所錄製之檔案音訊中，以 Praat 聲學分析軟體測量而得。該錄音原存放於國立臺灣大學人類學博物館，經由東京外國語大學語言資料庫取得。所測量之共振峰值透過對 eSpeak-NG 1.52 之 SPECTSQ2 二進位格式進行逆向工程，並開發 Python 工具，實作為可運作的語音合成器。與先前以轉寫符號近似選取之音素檔案相比，差異相當顯著——尤其是 /o/ (F1 : +80.5%) 及 /a/ (F2 : +30.5%) ——證明以轉寫為基礎的重新合成無法可靠呈現原始錄音中所記錄之音響特性。合成母音經語譜圖分析驗證，本研究方法論可作為以音響證據為基礎之消滅語言重建的通用框架。

*關鍵詞：巴賽語、福爾摩沙語族、聲學音韻學、共振峰測量、Praat、eSpeak-NG、檔案錄音、消滅語言文獻記錄*

## 一、緒論

巴賽語 (Basay, 亦拼作 Basai) 是一種已消滅的南島語系語言, 歷史上分布於臺灣東北部沿海地區, 即今日臺北、基隆及宜蘭平原一帶。就語言系屬而言, 巴賽語屬南島語系福爾摩沙語族, 與噶瑪蘭語 (Kavalan) 關係最為密切, 二者同屬東北福爾摩沙語支, 且為該語支僅有的兩個成員。巴賽語最後一批流利母語者於 1930 年代由日本語言學家淺井惠倫記錄 (Asai 1937), 其田野調查筆記及留聲機錄音構成本語言最主要且無可取代的文獻記錄。

自淺井的記錄以來, 後續研究——包括李壬癸 (Li 1993, 1999, 2001) 的奠基性研究及曾立洋 (Tseng 2022) 的比較重建研究——主要聚焦於音韻轉寫、語法分析與詞彙重建。這些研究成果彌足珍貴, 然而均未充分運用淺井檔案錄音中所蘊含的根本性資源: 音響訊號本身。就目前所知, 先前尚無研究直接從這批錄音中提取共振峰測量值, 使得巴賽語母音的音響特性迄今未獲量化。

本研究填補此一空缺。我們使用 Praat 聲學分析軟體, 直接從淺井針對特羅比灣方言說話者的錄音中測量 F1、F2、F3 共振峰頻率。這些以實證為基礎的數值, 進而透過對 eSpeak-NG 1.52 引擎的 SPECTSQ2 二進位格式進行逆向工程, 並開發 Python 工具, 實作為可運作的文字轉語音合成器。其成果是巴賽語母音的首個以音韻學為基礎的音響模型。

本研究同時證明先前方法——以轉寫符號相似性選取 eSpeak 音素檔案——的不足, 方式是將這些檔案的共振峰值與我們的實測目標值進行比較。差異在音響上相當顯著, 顯示以轉寫為基礎的近似法無法替代以測量為基礎的重建。

## 二、背景與先行研究

### 2.1 淺井惠倫錄音

本研究的主要音響資源, 是淺井惠倫於臺灣田野調查期間 (約 1936 年) 錄製的一批

留聲機錄音。錄音記錄了宜蘭地區巴賽語特羅比灣方言說話者的語音。原始錄音存放於國立臺灣大學人類學博物館（Museum of Anthropology, National Taiwan University），本研究透過東京外國語大學語言資料庫（Tokyo University of Foreign Studies Language Archive）取得錄音複本。錄音內容包括敘事文本，其中有「09-Ngazi 的故事（3）」（09-Ngazi 的故事（3））及「Mutravai 的故事（1）」等篇目。

特羅比灣方言代表巴賽語兩個已記錄方言區之一。淺井（1937）指出，其主要的特羅比灣方言受訪者 Mutravai 亦通噶瑪蘭語，不排除存在若干語音轉移的可能性。這一語言接觸情況被認定為音響資料代表性的局限，但不減損這批錄音作為巴賽語語音現存唯一音響記錄之重要性。

## 2.2 先行語言文獻記錄

淺井惠倫 1937 年的田野調查筆記提供了巴賽語音韻學最主要的轉寫記錄。李壬癸（1999）迄今提供最完整的音韻分析，確認子音音素約 15 個，母音系統含 4 至 6 個母音。曾立洋（2022）基於巴賽語與噶瑪蘭語的比較，提出原始東北福爾摩沙語的重建，含四個母音（\*i、\*u、\*ə、\*a）及四個雙母音。

上述研究均未嘗試對母音共振峰進行聲學測量。本研究直接以檔案錄音為基礎，提供此類測量值，而非依賴轉寫推論。

## 2.3 eSpeak-NG 與共振峰合成

eSpeak-NG 是一款開源文字轉語音引擎，採用以 Klatt 模型為基礎的共振峰合成器。母音音質透過 SPECTSQ2 格式的二進位檔案指定，其中編碼了含共振峰頻率、帶寬及振幅參數的頻譜幀。合成引擎直接從這些二進位檔案讀取共振峰值（peaks[1]、peaks[2]、peaks[3] 分別對應 F1、F2、F3），允許以赫茲（Hz）級精度指定共振峰。本研究開發了一種 Python 程式，直接從測量得的共振峰值生成 SPECTSQ2 格式檔案，

實現可重現、獨立於平台的合成流程。

### 三、研究方法

#### 3.1 音響測量

共振峰頻率使用 Praat (Boersma & Weenink) 從淺井檔案錄音中測量，分析對象為特羅比灣方言錄音，具體為敘事文本「09-Ngazi 的故事 (3)」及「Mutravai 的故事 (1)」。在語譜圖視圖中識別母音語料的穩態部分，並使用 Praat 共振峰追蹤器於各語料中點提取 F1、F2、F3 值。所得測量值代表淺井約 1936 年所錄製之特羅比灣方言說話者的巴賽語母音音響實現。第三段錄音「12-Mutravai 的故事 (4)」另行用於自動母音分類程序的方法論開發，未對合成器所實作之共振峰值有所貢獻。

為評估共振峰測量的可靠性並檢驗錄音雜音的潛在影響，本研究另行分析了第四段錄音：「13-Saturai 的故事 (1)」，這是一段歌唱錄音，共提取 9,991 個分析幀。以三個音源——用於實作的兩段散文敘事及該歌唱錄音——進行比較，六個母音的 F2 變動係數 (CV) 均低於 10% (範圍：1.9-6.0%)，四個母音 (/e/、/a/、/o/、/ə/) 的 F1 變動係數亦低於 5%。F1 變動係數較高的兩個母音 (/i/：16.3%，/u/：12.2%) 在散文錄音中的語料數量較少 (分別為 N=40 與 N=174)，顯示此係統計抽樣效應所致，而非音響測量誤差。上述結果表明，共振峰測量在不同錄音場次與語音風格之間均具有穩健性，1930 年代的錄音雜音對測量值並無實質影響。

帶寬估計值 (B1、B2、B3) 依據與測量共振峰配置相符之標準模態發聲規範設定，因為帶寬資訊無法從檔案錄音品質中可靠提取。測量共振峰值與帶寬估計值如表 1 所示。

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

表1. 巴賽語母音共振峰頻率 (特羅比灣方言)。F1-F3 係使用Praat 從淺井檔案錄音 (約1936年) 測量。帶寬 (B1-B3) 為模態發聲標準估計值。

### 3.2 SPECTSQ2 格式逆向工程

透過分析 eSpeak-NG 原始碼 (spect.cpp、spect.h、synthesize.h)，並與現有母音檔案的十六進位傾印進行交叉驗證，完成 SPECTSQ2 二進位格式的逆向工程。主要發現包括：

- 檔案標頭使用魔術簽章 SPECTSQ2 (0x53504543 + 0x54535132)，後接小端序長度前綴名稱字串。
- 浮點數值 (時間、基頻、長度、dx) 以 80 位元擴展精度 (每個 10 位元組) 儲存，而非標準 IEEE 754 64 位元雙精度浮點數。
- F1、F2、F3 分別編碼於 peaks[1]、peaks[2]、peaks[3]；peaks[0] 編碼基頻區域。

- KLATT\_Kopen (klatt\_param[] 索引 5) 應設為 50，以確保乾淨的有聲模態合成而無摩擦雜訊。

### 3.3 二進位檔案生成

六個巴賽語母音的 SPECTSQ2 檔案由 Python 程式 (gen\_basay\_vowels\_v8.py) 生成，以現有 eSpeak-NG 母音檔案為模板進行修補。修補流程如下：(1) 將 peaks[1-3].pkfreq 及 formants[1-3].freq 替換為測量得的 F1-F3 值；(2) 將 peaks[0].pkheight 及 peaks[4-8].pkheight 設為零，以抑制非共振峰頻譜成分；(3) 將諧波頻譜陣列 (spect[]) 清零，消除從模板母音繼承的頻譜著色效應。KLATT\_Kopen 值及其他合成參數從模板保留，確保乾淨的有聲輸出。

## 四、研究結果

### 4.1 與先前轉寫近似值的比較

在進行本次音響測量之前，eSpeak 中的巴賽語母音係以依轉寫符號相似性選取的現有音素檔案進行近似 (如 /i/ → i\_fnt、/o/ → o\_mid)。表 2 比較這些近似檔案的共振峰值與從淺井錄音測量所得之數值。

母音	參數	先前近似值 (Hz)	實測值 (Hz)	差異 (Hz)	差異 (%)
/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%
	F3	2477	2180	-297	-12.0%

母音	參數	先前近似值 (Hz)	實測值 (Hz)	差異 (Hz)	差異 (%)
/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%

表2. 先前轉寫近似值與淺井檔案錄音實測值之比較。\*i\_fnt 使用較舊的 SPECTSEQ 格式，無關鍵幀高度資料。

/o/ 的差異最為顯著 ( F1 : +80.5% , F2 : +34.9% ) , 其次為 /ə/ ( F1 : +47.7% ) 及 /a/ ( F2 : +30.5% ) 。 /o/ 的 F1 差距 359 Hz 在音響上具有重要意義 : 近似檔案將該母音置於中元音範圍 ( 446 Hz ) , 而實測值 ( 805 Hz ) 則識別其為較高的中後元音。後舌母音 ( /o/、/u/ ) 的 F3 值在先前近似中系統性偏高 400 至 536 Hz , 係不當使用前舌母音模板所致。

/i/ 的情況最為極端 : 先前使用的近似檔案 ( i\_fnt ) 採用較舊的 SPECTSEQ 格式 , 不含關鍵幀高度資料 , 使得該母音在本次重建之前實際上處於音響未定義的狀態。巴賽語合成器中 /i/ 的現行實作 , 完全以淺井錄音的實測值為依據 ( F1 : 269 Hz , F2 : 1696 Hz , F3 : 2128 Hz ) , 並無任何有效的先前近似值可供比較。這一事實進一步揭示了轉寫近似法的根本不足 : 對於 /i/ 這個在類型學上最為典型且跨語言最為穩定的高前元音 , 先前的方法根本無法產生任何可用的音響輸出。

## 4.2 語譜圖驗證

合成母音透過 SoX 進行語譜圖分析驗證。語譜圖確認 F1 排序遵循預期的類型學層級  $F1(/a/) > F1(/o/) > F1(/e/) > F1(/u/) > F1(/i/)$  , 與普遍的母音高度維度一致。前後對

立亦正確呈現：/i/、/e/、/ə/ 的 F2 集中於 1700 Hz 附近，/o/ 及 /u/ 的 F2 則在 1200 Hz 附近，與測量值及 /o/ 和 /u/ 預期之後舌母音特性一致。

所有合成母音在 4 kHz 以上均存在殘餘寬帶噪音，分析顯示此係 Klatt 合成引擎有聲化模型的特性所致，而非共振峰指定的問題，不影響 F1-F3 實作的音響效度。

## 五、討論

本研究的核心貢獻在於方法論層面：我們證明，即便是音質欠佳的檔案錄音，從中進行聲學測量，所得共振峰值仍與以轉寫為基礎的近似法存在實質性且系統性的差異。表 2 所記錄的差異並非隨機偏差，而是反映了近似法的特定結構性缺陷：選用 i\_fnt 對應 /i/ 產生了一個無可用共振峰資料的檔案；選用 o\_mid 對應 /o/ 使 F1 偏低 359 Hz，錯誤呈現其高度；對 /o/ 及 /u/ 使用前舌母音模板，則系統性地高估 F3 達 400 至 536 Hz。

若非回溯原始錄音，這些缺陷將無從發現。本研究證明，淺井的檔案錄音雖非為聲學音韻學目的而採集，其訊號品質仍足以進行基於 Praat 的共振峰提取。此一發現超越巴賽語本身，具有更廣泛的意涵：它表明，其他已消滅福爾摩沙語族語言的檔案錄音——若確實存在——或同樣適合進行聲學測量。

本研究的測量存在若干限制。首先，被錄製的說話者（特羅比灣方言）亦通噶瑪蘭語，無法完全排除語音轉移的可能性。其次，確切的錄音日期（約 1936 年）尚不確定。第三，帶寬值並非從錄音中測量，而是依標準模態發聲規範估算。未來研究應嘗試直接測量帶寬，並在能識別適合錄音的情況下，將特羅比灣方言的共振峰值與新社（Xinzhe）方言進行比較。

為評估測量可靠性，本研究分析了三段獨立錄音：用於共振峰實作的兩段散文敘事（09-Ngazi 的故事（3）及 Mutravai 的故事（1））以及一段額外的歌謠錄音（13-Saturai（1），N=9,991 幀）。三個來源的變異係數（CV）在六個母音的 F2 值均低於

10% (範圍：1.9–6.0%)，四個母音 (/e/、/a/、/o/、/ə/) 的 F1 值亦低於 5%。兩個 F1 變異較高的母音 (/i/：16.3%、/u/：12.2%) 在散文錄音中的語料數較少 (分別為 N=40 及 N=174)，顯示其變動可能源於統計抽樣誤差而非錄音雜訊。這些結果表明，共振峰測量值在不同錄音環境與語音風格之間具有穩健性，錄音雜訊對測量值的影響不顯著。

另一個尚未實作的參數是詞彙重音。巴賽語的重音位置無法從現有錄音中以足夠的信心加以確定：檔案音訊無法透過 F0 上升、時長或振幅等音響相關特徵可靠識別重音音節。初步觀察顯示，雙音節詞呈末尾重音、三音節詞呈倒數第二音節重音的規律，與南島語系常見的重音模板相符，但這仍是假設而非經實證確認的規則。目前的實作採用 eSpeak-NG 的預設重音指派。此外，經原始碼檢查確認，eSpeak-NG 1.52 版本的語音設定檔並未實作「stressrule」參數，因此重音指派無法單純透過 lang 檔案加以控制。待更可靠的音響證據出現後——最理想的方式是對檔案錄音中相同詞彙的多個語料進行時長與 F0 的系統性測量——應對此參數進行修訂。

本文所述的 eSpeak-NG 實作具有完整的可重現性。Python 生成工具、SPECTSQ2 格式文件及共振峰目標值均開放供研究社群使用，且隨著新音響證據的出現，合成器可隨時更新。

本研究一項特別值得關注的發現，涉及 /ə/ 與 /a/ 及 /o/ 在實測共振峰值與合成輸出中的近似融合現象。六個母音的 F2 值系統性地壓縮於狹窄的後方範圍 (F2：1191–1703 Hz)，其中 /a/、/o/、/ə/ 三個母音的 F2 尤為接近。這種壓縮現象與至少三個相互印證的因素相符。其一，高齡說話者的音韻性質漂移有充分文獻記錄，包括前舌及中央母音的後移、咽腔擴大、以及唇部收縮力減弱——這些因素均會縮減母音空間的 F2 跨距。其二，被錄製的說話者亦通噶瑪蘭語，而噶瑪蘭語重建的母音目錄 (\*i、\*u、\*ə、\*a) 並不包含獨立的 /o/ 與 /e/，可能導致該說話者的 /ə/ 與相鄰母音之間的音位對比減弱。其

三，曾立洋（2022）提出的原始巴賽語重建僅含四個母音（\*i、\*u、\*ə、\*a），顯示有聲巴賽語的 /e/ 與 /o/ 或許代表後期擴展或接觸借入，其音位強化程度相應較弱。因此，合成音中 /ə/ 與 /a/ 及 /o/ 難以知覺區辨，或忠實呈現了原始錄音中音位對比確實減弱的狀況，而非合成方法之局限。這一觀察本身即構成一項音韻發現：這批檔案錄音保存了母音空間收縮過程中的語音證據，與其他消滅語言及瀕危語言末期說話者語音學中所記錄的現象一致。

## 六、結論

本研究提出以直接從淺井惠倫檔案錄音（約 1936 年，特羅比灣方言，經由東京外國語大學取得）進行共振峰測量為基礎之巴賽語母音系統首次音響重建。六個母音的共振峰頻率 F1、F2、F3 以 Praat 提取，並透過逆向工程的二進位格式及 Python 合成工具在 eSpeak-NG 1.52 中實作。與先前轉寫近似值的比較揭示出實質性差異，若非回溯原始錄音，這些差異根本無從識別。所得合成器構成巴賽語首個以音韻學為基礎的音響模型，而本研究方法論為其他擁有現存檔案錄音之消滅語言的以證據為基礎之音響重建提供了可供參考的範本。

## 參考文獻

淺井惠倫（Asai, Erin）。1937。《バサイ》田野調查筆記。〔錄音約 1936 年，特羅比灣方言。〕原存放於國立臺灣大學人類學博物館，複本存於東京外國語大學語言資料庫。

Boersma, Paul & David Weenink. Praat: doing phonetics by computer.  
[www.praat.org](http://www.praat.org)

李壬癸（Li, Paul Jen-kuei）。1993。三種臺灣消滅語言的新資料。《中央研究院歷史語言研究所集刊》63(2): 301-323。

李壬癸 ( Li, Paul Jen-kuei ) 。 1999。 巴賽語若干問題。 收於 E. Zeitoun 及 P.J.K. Li  
（ 編 ） ， 《 第八屆南島語言學國際研討會論文選集 》 ， 635-664 頁。 臺北：中央  
研究院。

李壬癸 ( Li, Paul Jen-kuei ) 。 2001。 巴賽語的語言地位。 《 語言暨語言學 》 2(2):  
155-171。

曾立洋 ( Tseng, Li-yang ) 。 2022。 東北福爾摩沙語的重建。 碩士論文。

Dunn, Reece H. et al. eSpeak-NG. <https://github.com/espeak-ng/espeak-ng>