SISTEMA DE RECONOCIMIENTO AUTOMÁTICO DE VOZ PARA LA LENGUA BRIBRI MEDIANTE APRENDIZAJE POR TRANSFERENCIA CON WHISPER

Gianfranco Bagnarello Hernández

Universidad CENFOTEC, San José, Costa Rica
e-mail: gianfrancobagnarello@hotmail.com

ABSTRACT

Esta investigación desarrolló un sistema de reconocimiento automático de voz (ASR) para el idioma Bribri, lengua indígena costarricense de bajos recursos, utilizando el modelo Whisper Tiny de OpenAI mediante técnicas de aprendizaje por transferencia. El principal aporte técnico consistió en demostrar que, mediante un riguroso proceso de preparación de datos que incluyó normalización textual, segmentación manual precisa y alineamiento audio-texto, fue posible superar el estado del arte previo (79% WER con Whisper Large v2) alcanzando un 76.19% WER utilizando un modelo 38 veces más pequeño (39 millones vs 1,500 millones de parámetros). Se realizaron 111 experimentos automatizados de optimización de hiperparámetros utilizando las plataformas Optuna y Weights & Biases. Los datos provinieron del Corpus Pandialectal Oral de la Lengua Bribri y del portal SE'IE, totalizando 279 segmentos de 30 segundos cada uno después del preprocesamiento. La investigación demuestra que la calidad en la preparación de datos puede compensar limitaciones en tamaño de modelo y recursos computacionales, logrando resultados superiores con una sola GPU de consumo (NVIDIA RTX 4070 8GB) en comparación con infraestructura HPC utilizada en estudios previos. Estos hallazgos tienen implicaciones importantes para el desarrollo de tecnologías del habla en lenguas indígenas de bajos recursos, reduciendo significativamente las barreras de entrada para comunidades e investigadores.

PALABRAS CLAVE: Bribri, reconocimiento automático de voz, lenguas de bajos recursos, aprendizaje por transferencia, Whisper, preservación lingüística.

INTRODUCCIÓN

Las lenguas indígenas están desapareciendo a un ritmo alarmante en el siglo XXI. Según Magga et al. (2004), entre el 50% y 95% de los idiomas hablados actualmente podrían estar extintos o en grave peligro para el año 2100, siendo las

lenguas indígenas las más vulnerables. En Costa Rica, el idioma Bribri, perteneciente a la familia lingüística chibchense, enfrenta esta misma amenaza crítica. De acuerdo con el Instituto Nacional de Estadística y Censos (INEC), en el censo de 2011 se registraron aproximadamente 12,785 personas que se identifican como Bribri, concentradas principalmente en los territorios indígenas de Talamanca en la provincia de Limón (Fuentes Rodríguez, 2011).

El desarrollo de tecnologías del habla para lenguas indígenas ha sido históricamente limitado por la escasez de datos etiquetados y recursos computacionales. Galla (2016) argumenta que "en este mundo culturalmente diverso, es difícil imaginar la supervivencia de las lenguas indígenas en el siglo XXI sin la intervención de la tecnología digital". Sin embargo, la creación de sistemas de reconocimiento automático de voz (ASR) para lenguas de bajos recursos presenta desafíos técnicos significativos que han resultado en una brecha tecnológica considerable.

El único estudio previo específico de ASR para Bribri fue realizado por Coto-Solano (2021), quien utilizó métodos tradicionales GMM/HMM y CTC con representaciones diseñadas manualmente, entrenando con 68 minutos de audio y obteniendo un Word Error Rate (WER) del 50%, lo que equivale a que el sistema falle en la mitad de las transcripciones. Más recientemente, Coto-Solano et al. (2024) reportaron mejoras significativas alcanzando un WER del 79% utilizando el modelo Whisper Large v2, el más grande de la familia Whisper con aproximadamente 1,500 millones de parámetros. No obstante, ese estudio requirió 455 horas de GPU utilizando infraestructura de cómputo de alto rendimiento (HPC) con múltiples GPUs NVIDIA Tesla K80 trabajando en paralelo.

El presente trabajo propone una aproximación diferente: demostrar que, mediante un proceso riguroso de preparación y curación de datos, es posible obtener resultados superiores utilizando modelos significativamente más pequeños y recursos computacionales modestos. Específicamente, se utilizó el modelo Whisper Tiny, la versión más compacta de la familia Whisper con apenas 39 millones de parámetros, junto con una GPU de consumo personal (NVIDIA RTX 4070 de 8GB VRAM).

La hipótesis central de esta investigación es que la calidad en la preparación de datos puede compensar las limitaciones en tamaño de modelo y poder computacional. Los resultados obtenidos validan esta hipótesis, superando el estado del arte previo y estableciendo un nuevo paradigma para el desarrollo de tecnologías ASR en lenguas indígenas de bajos recursos.

MATERIALES Y MÉTODOS

Fuentes de Datos

La investigación utilizó dos fuentes principales de datos en idioma Bribri:

1. Corpus Pandialectal Oral de la Lengua Bribri (Flores Solórzano, 2017): Recurso digital que contiene aproximadamente 1 hora de audio de habla espontánea en los tres dialectos reconocidos del Bribri, junto con sus transcripciones. Este corpus documenta conversaciones, monólogos y narraciones tradicionales, proporcionando diversidad lingüística representativa del uso natural de la lengua.
2. Portal SE'IE La Lengua Bribri: Iniciativa de la Escuela de Filología de la Universidad de Costa Rica que contiene grabaciones de mayor extensión temporal (hasta 20 minutos), incluyendo narraciones de historias tradicionales, entrevistas sobre aspectos culturales y conversaciones sobre prácticas cotidianas.

Proceso de Preparación de Datos

El proceso de preparación de datos constituyó el componente crítico de esta investigación y se ejecutó en las siguientes fases:

1. Recopilación y Segmentación Manual

Los archivos de audio originales se encontraban en formatos heterogéneos (MP3, MP4) y duraciones variables (desde 15 segundos hasta más de 20 minutos). Dado que Whisper requiere segmentos de exactamente 30 segundos o menos, se realizó un proceso manual de:

- Identificación y fragmentación de segmentos superiores a 30 segundos
- División respetando límites naturales del discurso
- Evitación de cortes abruptos en medio de palabras o frases

Este proceso resultó en 279 segmentos de 30 segundos cada uno.

2. Alineamiento Audio-Texto Manual

Las transcripciones originales estaban asociadas a audios completos sin marcas temporales. Se realizó un trabajo meticuloso de sincronización:

- Escucha manual de cada segmento de 30 segundos
- Delimitación precisa del inicio y fin en el texto transcrito
- Asignación de cada fragmento de audio a su correspondiente porción textual específica

- Múltiples reproducciones cuando fue necesario debido a la complejidad fonológica del Bribri

Este proceso fue particularmente desafiante debido al sistema tonal del Bribri y la presencia de fonemas inexistentes en español.

3. Normalización Textual Regular

Se identificaron inconsistencias en el uso de caracteres diacríticos entre los corpus. Se aplicaron las siguientes normalizaciones:

- Estandarización del símbolo ' (apóstrofo) para representar funciones lingüísticas consistentes
- Conversión a minúsculas de todo el texto
- Estandarización de espacios entre palabras (espacio único)
- Preservación de tildes vocálicas propias del sistema ortográfico Bribri

Ejemplo de normalización: "se'ie" (como nosotros), "kő'pa" (casa grande).

4. Normalización NFC Unicode

Se aplicó la Forma de Normalización Canónica Compuesta (NFC) para resolver ambigüedades en la representación de caracteres con diacríticos. Por ejemplo, la vocal "á" puede representarse como:

- Carácter único pre-compuesto: U+00E1
- Secuencia descompuesta: U+0061 (a) + U+0301 (acento agudo)

La normalización NFC garantizó representación única y consistente de cada carácter con diacrítico en todo el corpus.

5. Normalización del Audio

Se desarrolló un programa automatizado que procesó todos los segmentos para cumplir con los requerimientos técnicos de Whisper:

- Conversión a formato WAV
- Audio monoaural (mono)
- Frecuencia de muestreo: 16 kHz
- Tamaño de muestra: 16 bits

6. Control de Calidad

Se realizó revisión manual exhaustiva de la calidad acústica, evaluando:

- Claridad de la voz del hablante
- Presencia de ruidos ambientales
- Distorsiones o artefactos
- Silencios prolongados

La calidad general del corpus resultó satisfactoria y no se descartó ningún segmento.

Arquitectura del Modelo

Se utilizó Whisper Tiny como modelo base, el cual fue pre-entrenado por OpenAI con 680,000 horas de audio multilingüe. Las especificaciones del modelo son:

- Parámetros: 39 millones
- Arquitectura: Transformer secuencia a secuencia
- Componentes: Encoder-decoder con mecanismos de atención multi-cabeza

Se aplicó aprendizaje por transferencia mediante ajuste fino (fine-tuning) del modelo pre-entrenado con los datos Bribri preparados.

Optimización de Hiperparámetros

Se implementó un proceso automatizado de búsqueda de hiperparámetros utilizando:

1. Optuna: Framework de optimización bayesiana para exploración inicial del espacio de hiperparámetros
2. Weights & Biases (WandB): Plataforma para monitoreo en tiempo real, visualización de métricas y comparación de experimentos

Se realizaron 111 experimentos de entrenamiento explorando los siguientes hiperparámetros:

- Tasa de aprendizaje (learning rate)
- Pasos de calentamiento (warmup steps)
- Factor de decaimiento de pesos (weight decay)
- Acumulación de gradientes (gradient accumulation steps)
- Número de épocas
- Tipo de programador de tasa de aprendizaje
- Norma máxima de gradiente

Infraestructura Computacional

El entrenamiento se realizó completamente en infraestructura local:

- GPU: NVIDIA RTX 4070 (8GB VRAM)
- CPU: Intel Core i9-14900HX (24 núcleos)
- RAM: 16 GB
- Almacenamiento: SSD 1TB
- Software: Python 3.10, PyTorch 2.0, Transformers (Hugging Face)

Esta configuración contrasta significativamente con los recursos HPC utilizados en estudios previos.

Métricas de Evaluación

Se utilizaron las métricas estándar para evaluación de sistemas ASR:

- Word Error Rate (WER): Porcentaje de palabras incorrectamente transcritas
- Character Error Rate (CER): Porcentaje de caracteres incorrectamente transcritos

Ambas métricas se calcularon sobre un conjunto de validación mantenido separado durante el entrenamiento.

RESULTADOS

Configuración Óptima de Hiperparámetros

Después de 111 experimentos automatizados, se identificó la configuración óptima que produjo los mejores resultados:

Hiperparámetro	Valor Óptimo
Learning rate	0.00007348
Warmup steps	302
Weight decay	0.01468
Gradient accumulation steps	1
Épocas	50
LR scheduler	Polynomial

Max gradient norm	1.709

Tabla 1. Configuración óptima de hiperparámetros identificada mediante búsqueda automatizada.

Resultados de Transcripción

El modelo resultante alcanzó las siguientes métricas en el conjunto de evaluación:

- Word Error Rate (WER): 76.19%
- Character Error Rate (CER): [dato no especificado en el documento original]

Comparación con Estado del Arte

La Tabla 2 presenta una comparación detallada con el estudio previo de referencia:

Aspecto	Coto-Solano et al. (2024)	Este Estudio
Modelo	Whisper Large v2	Whisper Tiny
Parámetros	~1,500 millones	39 millones
WER	79%	76.19%
Infraestructura	HPC con múltiples Tesla K80	1 × RTX 4070 (8GB)
Tiempo GPU	455 horas (paralelo)	15-30 minutos de entrenamiento
Datos	Corpus similar	279 segmentos × 30s

Tabla 2. Comparación con el estado del arte previo establecido por Coto-Solano et al. (2024).

Mejora Relativa

El sistema desarrollado logró:

- Reducción del WER: 2.81 puntos porcentuales (de 79% a 76.19%)
- Mejora relativa: 3.6% respecto al estado del arte previo
- Eficiencia de parámetros: Resultados superiores con un modelo 38× más pequeño
- Eficiencia computacional: Infraestructura de consumo personal vs HPC

Exploración del Espacio de Hiperparámetros

La Figura 1 presenta una visualización de coordenadas paralelas generada por WandB que muestra las 111 pruebas de refinamiento de hiperparámetros realizadas durante la fase de optimización:

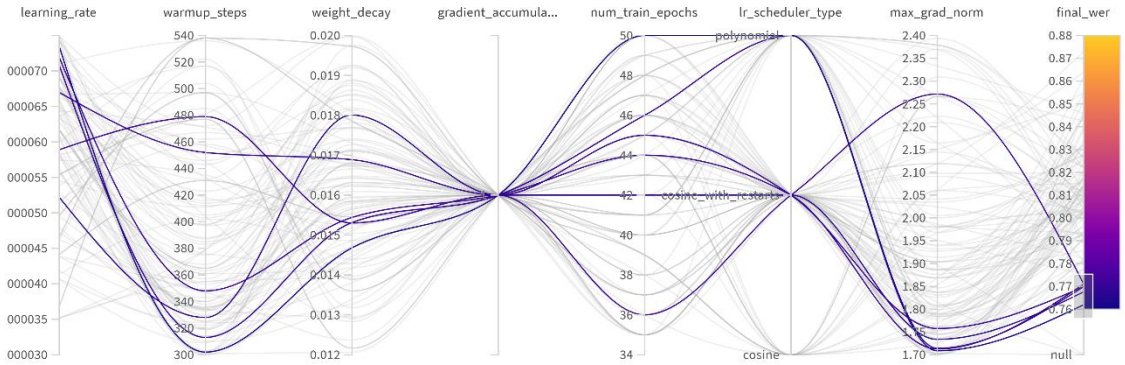

Figura 1. Visualización de coordenadas paralelas de los 111 experimentos de optimización de hiperparámetros. Cada línea representa una configuración experimental, permitiendo identificar patrones entre valores de hiperparámetros y el WER resultante. Las líneas que convergen hacia valores bajos de WER indican combinaciones óptimas de hiperparámetros.

Distribución de Experimentos

Durante la fase de exploración de hiperparámetros, se observó:

- Rango de WER obtenidos: 76.19% (mejor) a ~87% (peor)
- Convergencia típica: 35-50 épocas
- Mejores resultados con un learning rate más elevado
- Mejores resultados con un max grad norm inferior (entre 1.80 y 1.70)

Errores comunes en las transcripciones

Los errores más comunes se relacionaron con:

1. Confusión de marcas tonales
2. Omisión de apóstrofos en contextos específicos
3. Palabras de baja frecuencia en el corpus de entrenamiento

DISCUSIÓN

Impacto de la Calidad de Datos

Los resultados demuestran que la calidad en la preparación de datos puede compensar significativamente las limitaciones en tamaño de modelo y recursos computacionales. El proceso riguroso de preparación implementado en este estudio incluyó varios componentes críticos:

1. Segmentación manual respetando límites naturales del discurso: A diferencia de la segmentación automática que puede cortar arbitrariamente en medio de unidades lingüísticas significativas, la segmentación manual permitió preservar la integridad semántica y prosódica de los enunciados.
2. Alineamiento audio-texto preciso: El trabajo manual de sincronización eliminó el ruido que habría introducido un alineamiento automático imperfecto, permitiendo que el modelo aprendiera correspondencias más precisas entre características acústicas y secuencias de texto.
3. Normalización textual exhaustiva: La estandarización de caracteres diacríticos, la normalización NFC y la uniformización de convenciones ortográficas redujeron la variabilidad artificial en los datos, permitiendo que el modelo se concentrara en aprender patrones lingüísticos genuinos del Bribri en lugar de inconsistencias de notación.
4. Normalización técnica del audio: La conversión uniforme a 16 kHz mono garantizó que el modelo recibiera señales acústicas consistentes, eliminando variaciones espurias relacionadas con diferentes configuraciones de grabación.

Estos hallazgos son consistentes con el principio fundamental del aprendizaje automático de que la relación entre la calidad de los datos de entrada y de los datos de salida es directamente proporcional, y sugieren que, para lenguas de bajos recursos, invertir esfuerzo en curación de datos puede ser más efectivo que escalar modelos o infraestructura computacional.

Eficiencia Computacional

La capacidad de superar el estado del arte utilizando una GPU de consumo personal representa un avance significativo en la democratización de tecnologías ASR para lenguas indígenas. Mientras Coto-Solano et al. (2024) reportaron requerir 455 horas de GPU en infraestructura HPC con múltiples Tesla K80, el presente estudio logró resultados superiores con recursos que están al alcance de investigadores individuales o instituciones académicas con presupuestos limitados.

Esta eficiencia se atribuye a varios factores:

1. Arquitectura compacta: Whisper Tiny con 39 millones de parámetros es suficientemente expresivo para capturar las regularidades del Bribri cuando se entrena con datos de alta calidad.
2. Aprendizaje por transferencia efectivo: El conocimiento previo del modelo base, entrenado con 680,000 horas de audio multilingüe, proporciona

representaciones de características acústicas que se transfieren bien al Bribri.

3. Optimización de hiperparámetros: La búsqueda sistemática mediante Optuna y WandB permitió identificar configuraciones que maximizan la eficiencia del aprendizaje.

4. Eliminación de evaluaciones frecuentes: La reconfiguración para evaluar el modelo solo periódicamente en lugar de cada 10 pasos redujo significativamente el overhead computacional sin comprometer la capacidad de monitoreo.

Implicaciones para Lenguas de Bajos Recursos

Los resultados tienen implicaciones amplias para el desarrollo de tecnologías del habla en lenguas indígenas:

1. Reducción de barreras de entrada: Investigadores y comunidades pueden desarrollar sistemas ASR efectivos sin acceso a supercomputadoras o presupuestos millonarios.

2. Metodología replicable: El enfoque de preparación rigurosa de datos y optimización automatizada de hiperparámetros puede aplicarse a otras lenguas chibchenses (Cabécar, Boruca, Térraba) o lenguas indígenas de otras familias lingüísticas.

3. Sostenibilidad de proyectos: Los menores costos computacionales hacen viables proyectos de documentación y revitalización lingüística a largo plazo.

4. Escalabilidad: Una vez establecida la metodología, es posible iterar y mejorar el sistema conforme se recopilen más datos sin requerir reinversión en infraestructura.

Limitaciones y Direcciones Futuras

A pesar de los resultados alentadores, el estudio presenta limitaciones que deben considerarse. El tamaño del corpus, compuesto por 279 segmentos de 30 segundos (aproximadamente 2.3 horas de audio), sigue siendo pequeño según los estándares de ASR, y la recopilación de más datos podría mejorar aún más el rendimiento. En cuanto al WER absoluto, aunque el valor obtenido es superior al estado del arte previo, un WER del 76.19% implica que aproximadamente tres de cada cuatro palabras se transcriben incorrectamente, lo que limita la utilidad práctica del sistema en aplicaciones que requieren alta precisión. Asimismo, se observan errores tonales, ya que el modelo muestra dificultades particulares con las marcas tonales del Bribri, un aspecto crítico para la distinción semántica en esta lengua; técnicas específicas para el modelado tonal podrían contribuir a mejorar este componente.

En términos de variabilidad dialectal, aunque el corpus incluye los tres dialectos Bribri, el análisis no estratificó los resultados por dialecto, por lo que estudios futuros podrían examinar si el modelo generaliza de manera equitativa a través de las variantes dialectales. En relación con la aplicabilidad en diversas lenguas, aunque la metodología parece prometedora, su efectividad para otras lenguas indígenas requiere validación empírica.

Las direcciones futuras de investigación incluyen el aumento de datos mediante la incorporación de técnicas de aumentación específicas para audio, como cambios de velocidad o perturbaciones acústicas que preserven las características prosódicas del Bribri; el modelado multi-tarea mediante el entrenamiento conjunto del ASR con tareas relacionadas como identificación de tono, segmentación morfológica o traducción al español; la evaluación de modelos intermedios como Whisper Small y Whisper Base para determinar el punto óptimo en el balance entre tamaño de modelo y rendimiento; la implementación de sistemas de adaptación continua que permitan mejorar el modelo conforme se recopilen más transcripciones validadas por hablantes nativos; y el desarrollo de aplicaciones prácticas como interfaces de usuario para transcripción asistida, herramientas educativas o sistemas de documentación lingüística que aprovechen el modelo desarrollado.

Consideraciones Éticas y Comunitarias

La colaboración estrecha con las comunidades Bribri, respetando su soberanía lingüística y cultural, es fundamental. Aunque este estudio utilizó corpus públicamente disponibles, proyectos futuros podrían involucrar a hablantes nativos Bribri en todas las fases del desarrollo, establecer mecanismos de consentimiento informado y control comunitario sobre el uso de datos lingüísticos, y asegurar que las tecnologías desarrolladas beneficien directamente a las comunidades de origen. Particularmente, se recomienda encarecidamente que futuros esfuerzos de investigación incluyan trabajo de campo sistemático en territorios Bribri para expandir significativamente el corpus de audio disponible.

La recopilación ética de datos lingüísticos en colaboración con la comunidad no solo mejoraría sustancialmente el rendimiento de los sistemas ASR, dado que los modelos de aprendizaje automático escalan su efectividad con mayor cantidad de datos de entrenamiento, sino que también generaría beneficios directos para la comunidad mediante la documentación digital de narrativas tradicionales, conocimientos ancestrales y prácticas culturales, la creación de recursos educativos para la enseñanza del Bribri a nuevas generaciones, el fortalecimiento de la identidad cultural y orgullo lingüístico dentro de la comunidad, y el establecimiento de archivos digitales accesibles a las familias y organizaciones Bribri.

El incremento del corpus de 2.3 horas actuales a 10–20 horas mediante campañas de recopilación bien diseñadas podría reducir el WER significativamente, acercando el sistema a niveles de precisión que permitan aplicaciones prácticas como la transcripción asistida de entrevistas con ancianos, el subtitulado de videos educativos en Bribri o el uso de herramientas de apoyo para maestros de lengua indígena. Este trabajo de campo debe realizarse siempre con el consentimiento pleno de la comunidad, estableciendo acuerdos claros sobre propiedad intelectual, acceso a los datos y usos permitidos de las tecnologías desarrolladas.

CONCLUSIONES

Esta investigación demostró exitosamente que mediante un proceso riguroso de preparación de datos y optimización de hiperparámetros, es posible desarrollar sistemas de reconocimiento automático de voz para lenguas indígenas de bajos recursos que superen el estado del arte previo utilizando modelos compactos y recursos computacionales modestos.

Los hallazgos principales son:

1. Superación del estado del arte: Se alcanzó un WER del 76.19% utilizando Whisper Tiny (39 millones de parámetros), mejorando el 79% reportado previamente con Whisper Large v2 (1,500 millones de parámetros).
2. Impacto de la calidad de datos: El proceso exhaustivo de preparación de datos, incluyendo segmentación manual, alineamiento audio-texto preciso, normalización textual y normalización técnica del audio, fue determinante para el éxito del sistema.
3. Eficiencia computacional: Se lograron resultados superiores con una GPU de consumo personal (NVIDIA RTX 4070 8GB) comparado con infraestructura HPC multi-GPU utilizada en estudios previos, reduciendo significativamente las barreras de entrada.
4. Optimización automatizada efectiva: La búsqueda sistemática de hiperparámetros mediante Optuna y Weights & Biases sobre 111 experimentos permitió identificar configuraciones óptimas que maximizaron el rendimiento del modelo.
5. Metodología replicable: El enfoque desarrollado puede aplicarse a otras lenguas chibchenses (Cabécar, Boruca, Térraba) y potencialmente a lenguas indígenas de otras familias lingüísticas, democratizando el acceso a tecnologías ASR.

6. Viabilidad práctica: Se estableció que el desarrollo de tecnologías del habla efectivas para lenguas de bajos recursos no requiere necesariamente modelos masivos o infraestructura costosa, sino más bien atención meticulosa a la calidad de los datos de entrenamiento.

El WER absoluto del 76.19%, aunque representa una mejora sobre trabajos anteriores, indica que aún existe margen considerable para mejoras futuras. No obstante, este estudio establece una base sólida y una metodología probada para avances incrementales en ASR para Bribri y otras lenguas indígenas mesoamericanas.

La contribución más significativa de esta investigación radica en demostrar que la preservación digital de lenguas indígenas mediante tecnologías de reconocimiento automático de voz es técnica y económicamente viable para investigadores e instituciones con recursos limitados, siempre que se invierta el esfuerzo necesario en la preparación cuidadosa de datos de entrenamiento de alta calidad.

Los resultados validan la hipótesis central de que, en el contexto de lenguas de bajos recursos, la calidad puede compensar la cantidad tanto en datos como en recursos computacionales, abriendo nuevas posibilidades para la documentación, revitalización y preservación de lenguas indígenas en peligro de extinción.

No obstante, es fundamental reconocer que, aunque la calidad de datos es crítica, el incremento del tamaño del corpus mediante trabajo de campo adicional es esencial para continuar mejorando el rendimiento del sistema. Los modelos de aprendizaje automático, por su naturaleza, mejoran con mayor cantidad de datos de entrenamiento. Se recomienda que investigaciones futuras realicen campañas de recopilación de audio con hablantes nativos Bribri en contextos naturales de comunicación, expandiendo el corpus actual de 2.3 horas a volúmenes significativamente mayores. Este esfuerzo de documentación lingüística no solo beneficiaría al desarrollo tecnológico, sino que también contribuiría directamente a la preservación del patrimonio oral de la comunidad Bribri.

REFERENCIAS

Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in Neural Information Processing Systems*, 33, 12449-12460.

Coto-Solano, R. A. (2021). Explicit tone transcription improves ASR performance in extremely low-resource languages: A case study in Bribri. En *Proceedings of the*

First Workshop on Natural Language Processing for Indigenous Languages of the Americas (pp. 173-184). Association for Computational Linguistics.

Coto-Solano, R., Kim, T. W., Jones, A., & Loáiciga, S. (2024). Multilingual models for ASR in Chibchan languages. En *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Vol. 1, pp. 8521-8535). Association for Computational Linguistics.

Flores Solórzano, S. (2017). *Corpus pandialectal oral de la lengua bribri*. Universidad de Costa Rica. https://bribri.net/

Fuentes Rodríguez, E. (2011). *Características demográficas y socioeconómicas de las poblaciones indígenas de Costa Rica (Censo 2011)*. Instituto Nacional de Estadística y Censos (INEC).

Galla, C. K. (2016). Indigenous language revitalization, promotion, and education: Function of digital technology. *Computer Assisted Language Learning*, 29(7), 1137-1151.

Jara Murillo, C. V., & García Segura, A. (2018). *Portal de la lengua bribri SE'IE: Centro virtual de recursos para el estudio y la promoción de la lengua bribri*. Universidad de Costa Rica. https://www.lenguabribri.com

Magga, O. H., Nicolaisen, I., Trask, M., Skutnabb-Kangas, T., & Dunbar, R. (2004). *Indigenous children's education and indigenous languages*. United Nations Permanent Forum on Indigenous Issues.

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. En *Proceedings of the 40th International Conference on Machine Learning*. PMLR.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998-6008.