International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Development of Speech Corpus for Improving Phonetic Search Engine Performance in Zero Recourse Nyishi Language

Likha Ganu[1*], Biri Arun [2]

Submitted: 14/03/2024 **Revised**: 28/04/2024 **Accepted**: 05/05/2024

Abstract: Advancements in speech translation technology are underway to enable natural communication across languages. Most languages with limited resources don't even have any speech data. Creating speech corpora is extremely difficult and time-consuming. This paper outlines our ongoing endeavor to construct speech corpora for one of the zero-recourse languages in North-East India, with a specific focus on the Nyishi Language from the Tibeto-Burman language family. Methodology involving lab-based and crowd-sourced recordings using handheld audio recorders, and the Nyishi speech corpus database currently boasts more than 2200 utterances from 34 native speakers diverse in regional dialect, age, and gender. The corpus encompasses four distinct speech modes - spontaneous conversations, fluent speech, read speech, and storytelling narratives. Audio recordings were meticulously transcribed using International Phonetic Alphabet symbols and annotated for tone, pitch, pacing, syllabification, and break marking. Statistical analysis of the phoneme distributions provides new insights into the phonetic composition of Nyishi. Findings reveal the central role of the vowel /a/ (34.1% instances) along with prolific use of front-vowel-based diphthongs like /ai/ (15.5% instances) and triphthong formations containing /ia/ sequences (30.2% instances). With speaker metadata encoded directly into the filename conventions, this structured corpus supports diverse research inquiries from acoustic phonetics, tonality studies, and morphological analysis to the development of speech recognition and synthesis systems. Constructing usable speech recognition and synthesis datasets for endangered languages like Nyishi facilitates preservation efforts and enables language revitalization applications. This paper elaborates on the methodology employed in collecting speech samples and presents descriptive statistics of the speech corpora.

Keywords: *Speech Corpus, Phonetic engine, Nyishi, Transcription, Speech Recognition.*

1. Introduction

The Tibeto-Burman language family consists of around 400 languages spoken primarily in the Himalayan region. Many of these languages are considered low-resource, lacking sizable annotated speech corpora[1]. This poses challenges for developing speech technologies. Acoustic-phonetic analysis examines properties like pitch, formants, and duration to characterize speech sounds and phonemes. Nyishi is also one of the low-resource languages used by the Nyishi people in the Indian state of Arunachal Pradesh. The Nyishi language belongs to the Tani group of languages, which is a distinct subgroup within the Tibeto-Burman language family. Other languages within this language family include Adi, Bangni, Bokar, Bori, Damu, Gaol, Hill Miri, Milang, Na, Tagin [2][3].

Besides the above-mentioned varieties, the Nyishi language has been provisionally matched up with some other Tibeto-Burman languages like that of Bhutan Tshangla, Written Burmese, Written Tibetan, Proto-Tani, Pro Researchers have made these comparisons to help them understand the linguistic features of Nyishi and how it relates to other languages. To date, there is little research which was done on The Nyishi Language specially, making this language be understudied in the study areas of linguistics and speech processing. Consequently, Nyishi should be documented to secure its linguistic heritage and also for the sake of Tibeto-Burman language family studies. With the dynamic world we live in today, the importance of correct documenting and preserving low-resource languages such as Nyishi cannot be underestimated. Collection of speech data, therefore, forms the most critical step in building the Nyishi corpus of speech. We have used credible sources that relate to the linguistic affinity and features of the Tibeto-Burman language family, including the Tani group to which Nyishi belongs [4].

Although there are few studies on Nyishi language, this is attributed to the shortage of resources and the lack of a recognized script used for writing this language.A speech corpus is necessary for phonetic analysis, speech synthesis development, and linguistic research and preservation due to the above noted challenges. The resultant speech corpus would be an important tool for linguists in their study of Nyishi, or even simply a record of Nyishi speech that could potentially help

[1*,2]*Department of Computer Science and Engineering, National Institute of Technology Arunachal Pradesh, Jote.National Institute of Technology Arunachal Pradesh, Jote, Itanagar, Papumpare,791113, Arunachal Pradesh, India.* (http://orcid.org/0000-0003-3695-9195)[1*], (http://orcid.org/ 0000-0002-8591-3989)[2]
***Corresponding Author:** Likha Ganu
**E-mail: ganu664@gmail.com*

maintain the language. Overall, it is crucial to build a and preservation of the low-resource Nyishi language and the development of speech recognition and synthesis technologies specific to the Nyishi language. In this study, we provide an overview of the Nyishi speech recording project. The primary goal of this study is to develop a speech corpus for the Nyishi language, spoken in various regions of Arunachal Pradesh, including Kradadi, Kurung Kumey, Lower Subansiri, Papumpare, and some parts of Kamle district shown in Fig 1. For Nyishi speech corpus for accurate documentation fundamental studies of phonetic, prosodic, and syntactic structures characteristic to Nyishi language will be the first corpus of great value for Nyishi linguists. Also, this research will be essential for saving endangered Nyishi language, developing tools for automatic speech recognition, and Therefore, we performed speech recording among native Nyishi speakers in different age and social groups.

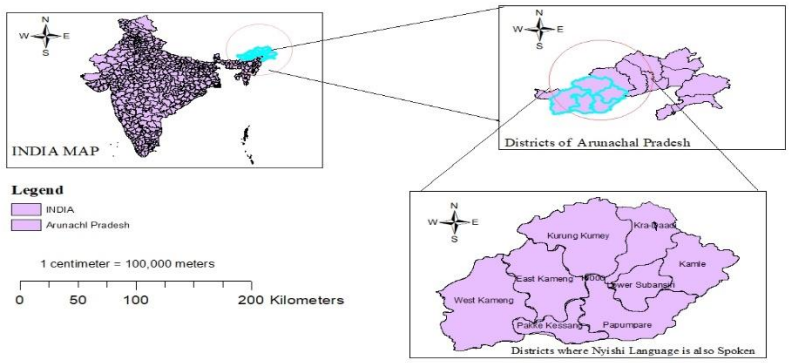

Fig.1 District of Arunachal Pradesh where Nyishi is spoken

2. Speech Corpus

Speech corpus is the collection of speech samples which are utilised for creation and assessment of several speech technologies like Speech Synthesis technology and Automatic Speech Recognition [5]. Therefore, large databases of voice audio recordings and transcripts, which are also known as speech corpora, play an important role in this endeavor. These are data intensive technologies which involve training algorithms and models to successfully interpret spoken language using appropriate recognition mechanisms. The present literature review is to investigate contemporary studies on speech corpus development in low-resource languages and its applications. There is also a corpus known as the ALFFA speech corpus [6], which consists of recordings in eight different African languages such as Amharic, Hausa, Yoruba, Wolof, Swahili, Zulu, Sesotho sa Leboa, and The data was collected by prompting native speakers to read predefined sentences from a phone or computer. In total, the corpus contains 16-23 hours of speech data per language from 76-125 speakers. The corpus has a training-test split and includes parallel translations in English. As noted by the authors, ALFFA is the largest publicly available dataset for these African languages to date.

The MLVS-3 corpus [7] consists of telephone speech data in 6 languages: Assamese, Bengali, Gujarati, Kannada, Malayalam and Telugu. The speech data was obtained from anonymized voice call recordings with consent from native speakers who were prompted to read predefined passages. The corpus contains 30 hours of speech per language with associated transcripts and glossary details. The diverse recording conditions and speakers make this dataset suitable for building robust speech models.

However, building a speech corpus involves careful consideration of the data collection methodology. The quality of record, the diverse speaker and the size of the corpus can affect how useful the end product will be [8]. Two predominant approaches have emerged for speech data collection: lab-based recording and crowdsource [9]. Several approaches have been proposed. The standard lab-based approach entails having individuals invited into a recording studio where they will have to recite scripts or undertake guided undertakings. The strategy ensures quality in terms of regulation of recording procedures as well as the description of speakers. Nonetheless, it is expensive and laborious to enroll and record many speakers in this manner. Crowdsourcing has become popular as an efficient alternative for large-scale speech data collection. Services like Amazon Mechanical Turk allow recruiting and recording speakers remotely via internet-connected devices [12]. While this approach is more scalable, the

recordings can vary greatly in quality due to uncontrolled recording environments. Well-known crowdsourced corpora include VoxForge [11] and LibriSpeech [10]. Hybrid approaches attempt to balance the benefits of both lab and crowdsourced recording. Some researchers send audio recording kits to participants' homes to collect high-quality samples remotely [13]. Others use a wizard-of-oz methodology where remote participants believe they are interacting with an AI system, but a hidden human wizard elicits naturalistic speech [14]. Data augmentation techniques like speed perturbation and volume perturbation can increase small datasets. Voice cloning using neural networks is another augmentation strategy [15]. However, while augmentation helps, it cannot replace real diverse speech data. Other work has focused on efficient data collection methods.

This study to Build a speech corpus for Nyishi would involve Hybrid approaches both traditional method and crowd-sourced recording because of limited resources, participants are brought into a recording studio and prompted to read scripts or engage in directed tasks method to collect and transcribe a wide range of spoken data from native speakers of the language. This Nyishi speech corpus should be able to provide enough data to extract the distinct phonetic, prosodic, and syntactic features of the language. Moreover, the corpus would be useful in describing language personality characteristics like characteristic speech patterns, words, grammar and discourse markers for Nyishi. It would also help in conservation and writing history of Nyishi culture and language. The spoken language collected resources will also be accompanied by translations from a high resource language such as English and IPA transcriptions. This improves the usability of the corpus and allows even non-linguistics experts enjoy it.

We have studied the 4 key areas of novelty in the research presented in this paper on developing a Nyishi speech corpus:

- **First speech corpus for the low-resource Nyishi language:** This paper documents the first efforts towards building an annotated speech dataset for Nyishi, a critically endangered Tibeto-Burman language lacking in computational linguistic resources. Constructing such a corpus can enable future speech technology development and documentation.
- **Analysis of Nyishi phonetics and phonology:** The corpus provides data enabling phonetic and phonological analysis of Nyishi's tone patterns, some of the vowels and consonant inventories, syllable structure tendencies, and prominence of diphthongs/triphthongs - unlocking insights into this understudied language's sound system.
- **Speaker diversity:** 34 Nyishi speakers across genders, regional backgrounds, and wide age ranges contributed speech samples, capturing different dialects. This diversity can facilitate the study of linguistic variation and multi-genre speech collection: Speech was gathered through lab sessions, crowdsourcing, and on-location district-level recordings across spontaneous conversations, fluent narratives, and read speech.
- **IPA-based transcription methodology on unstudied counting genres** - highlighting Nyishi language use across communicative contexts: Meticulous phonetic transcription of Nyishi utilized International Phonetic Alphabet symbols annotated for tone, syllabic constituents, pitch, pacing, and phrases. This enables precise documentation of Nyishi's phonetics and sound patterns.

In this study, we contribute to initial phonetic-acoustic documentation and the first corpus-based data for Nyishi, one of the significantly impoverished languages. The construction of this corpus will be of crucial significance for preservation programmes, dialectology investigations, computational models, and exploration for distinctive linguistic structures in Bodo-Nyishi within the Tibeto-Burman group.

3. METHODOLOGY

Unlike in other big languages such as English and Mandarin Chinese, Nyishi has fewer available resources. Some of these papers and dictionaries, for example, [16], [17]. Nevertheless, advanced computational resources such as text corpora, speech databases, and treebanks are unavailable. This therefore poses a challenge in developing natural language processing tools and applications for the Nyishi. This calls for Nyishi speech corpus development in Arunachal Pradesh has tremendous implications for a range of players. This will result in an extensive corpus of Nyishi language samples for the linguists and researchers to conduct a detailed phonetic, phonological and linguistic structure analysis. To track the language evolution and possible variations among different age groups of speakers, the speech of speakers' of different ages was recorded. The Nyishi speech corpus will be very significant in documenting the language for future generations in language preservation effort. In addition, the corpus will facilitate growth and development of language technologies such as the Nyishi automatic speech recognition and text-to-speech systems. The Nyishi community will therefore be able to harness and utilize speech-based technologies in spheres such as education, health and communication.

3.1. Text Corpora

The majority of the text corpora for the Nyishi language are folktales and translations from the Bible [20]. Larger corpora covering diverse domains are needed for training language models and other NLP applications. Speech corpora to support automatic speech recognition are also lacking. Developing annotated spoken and textual corpora is critical. The text corpora reported in this paper comprises text in two languages, English and Nyishi Language. In this preliminary effort, Nyishi's sentences were collected from different sources such as Bible translations, local story books, online platforms in which Nyishi's language is spoken, etc. In an ongoing effort, Nyishi's speech was transcribed into IPA Based transcription, and the same sentence in transcribed into English. Different mode of recording and transcription was done to enhance phonetic richness, i.e., to enhance the frequency of rare triphones. In addition to the sentences, proverbs and digit sequences were added to the Nyishi text corpora. Currently, the Nyishi text corpus contains 10020 unique sentences with a vocabulary of 14600 unique words. The 10020 sentences are arranged in more than 250 different sets. The sentences are segmented such that their length lies between 4 to 15 words.

3.2. Corpora File Naming

Speech corpora comprise large databases of text and audio recordings collected from multiple speakers. They are essential resources for enabling the automatic retrieval of linguistic, para-linguistic, and non-linguistic information from speech data and developing speech technology applications like automatic speech recognition systems. However, building high-quality corpora requires careful planning and consistent protocols during the data collection process. An important consideration is establishing conventions for naming the audio files that will maximize the organization, clarity, and utility of the corpus.[18] note that thoughtful file names allow users of the corpus to quickly parse key recording details like speaker identity, content, language, recording conditions, and date. We therefore use the file name convention that was adopted by a group of Indian institutes that worked together to create a speech recognition system for Indian languages [19]. To enable clear computer program parsing, the file name is formed as a series of codes that are alternatively encoded in terms of Roman alphabets and numerals. The naming convention for the Nyishi Speech corpus can be seen in Table.1 The Nyishi Speech corpus is being recorded in three different Modes of Speech: Spontaneous speech mode (1), Fluent speech mode (2), and Read mode (3). And gender will be denoted by a single letter M (male) and F (female). Three digit serial number read for Speaker ID (1XX to 999) and for recording device, Microphone (1), Digital Recorder zoom H6 (2), Mobile Phone(3), For age it is denoted with single letter (A) Range 10-18 years, (B)Range 19-40 years, (C) Range 41 years above. Language code is Nyishi (20) and English (00) and the last is the types of Target syllable conventions, especially for read mode recording, for Monosyllable (x), Disyllabic(y), Tri Syllable (z), Tetra syllabic(w) the naming convention is designed to facilitate the analysis of linguistic and phonetic attributes in speech, as well as to enable seamless parsing by various speech technology models.

Table 1. File naming Convention of Corpus

<table>
<tr><td colspan="4">MDTW (Mono , Dy ,Tri Syllabic Word)</td></tr>
<tr><td>Mode of Speech 1,2,3</td><td>4. Device</td><td colspan="2">6. Language</td></tr>
<tr><td></td><td>Microphone 1</td><td colspan="2">Nyishi 20
English 00</td></tr>
<tr><td>Gender</td><td>Digital Recorder (zoom H6) 2</td><td colspan="2">7. Phonemes Sounds Domain.
[Mono(x), Dy(y), Tri(z),Tetra(w) Syllables]</td></tr>
<tr><td>Male M</td><td>Mobile Phone 3</td><td>Sounds Categories Code</td><td></td></tr>
<tr><td rowspan="4">Female F</td><td rowspan="4"></td><td>a = 1</td><td>e = 2</td></tr>
<tr><td>i = 3</td><td>o = 4</td></tr>
<tr><td>u = 5</td><td>ə =6</td></tr>
<tr><td>New sound = #</td><td>ɨ =7</td></tr>
<tr><td></td><td>5. Age</td><td colspan="2"></td></tr>
<tr><td>Speaker ID XXX</td><td>Range 10-18 years A</td><td colspan="2"></td></tr>
<tr><td></td><td>Range 19-40 years B</td><td colspan="2"></td></tr>
<tr><td></td><td>Range 41 years above C</td><td colspan="2"></td></tr>
</table>

The example file name is 1MXX12B20x.wav

3.3. Speech Data Collection

Our data collection methodology focused on the native Nyishi speakers across 5 key districts - Kradaadi, Kurung Kumey, Lower Subansiri, Papumpare, and parts of Kamle - in Arunachal Pradesh where the language is actively spoken [20]. We traveled extensively through these regions to obtain a broad representation of the Nyishi dialects and speaker demographics. Considering that the Nyishi language has not previously been the subject of acoustic-phonetic research and lacks a codified orthography, our priority was eliciting high-quality recordings of vowel sounds. We strategically compiled a list of target vocabulary that would allow us to sufficiently identify the Nyishi vowel inventory and the rest of the phonemes. With different modes and with specialized wordlist, we then recorded a robust sample of 34 native speakers of both genders across a wide age range. 24 of the speakers were male, while 10 were female. Within the male group, 18 were young adults aged 18-30 years, 5 were middle-aged adults aged 31-50 years, and 1 male elderly speaker aged 51+ years. Of the female speakers, 6 were young adults aged 18-30 years while 4 were middle-aged adults aged 31-50 years. From this diverse pool of Nyishi dialects, ages, and genders, we obtained natural speech samples that contained multiple iterations of the critical vowel phonemes that our wordlist was designed to elicit. Every recording was made with the goal of improving speech synthesis technology and acoustic-phonetic analysis..

3.4. Recording mode

The approach to developing a speech corpus in the Nyishi language includes recordings in three different modes: Read mode, fluency mode, and speech mode. Spontaneous speech mode was adopted whereby such natural discussions and interactions among native speakers would be captured directly without any rehearsals or preparations. Therefore, this provides opportunity to collect authentic language use and expressions in natural settings. The final recordings involved speakers who had full command of the Nyishi language and speaking abilities. The fluent speech mode looks to capture fluent, cohesive speech [21]. The recordings enable an individual to understand the vocabulary and grammar of the Nyishi language being used where the speaker conveys complex messages using stories.The last is read mode, which involves recording speakers who are given written prompts or passages to read out loud. The read mode recordings involve scripted materials such as stories, poems, or traditional texts, which allow for a more controlled exploration of specific linguistic features and formal language usage (Corpora compilation for prosody-informed speech processing | Language ..., n.d). For example, a prior preparation or rehearsal was done for targeting specific phonemes or Monosyllabic, disyllabic, and Multisyllabic sounds of interest words to understand the syntax and semantics of the languages. Spoken speech was recorded in different modes designated for different purposes syllable structure of the speech can appear mostly in V, CV, VC, CVC, CCV, VCV, CCVC, CVVC, CVCCVC, syllable format in the sentence of speech in corpus.

The speakers were asked to utter each target word in a sentence, in isolation, and a carrier phrase. For example, if the target word for monosyllable is “Bag”, In Nyishi it is pronounced as ‘ʧuk’ the speakers will utter as

(a) ʧuk, ʧuk, ʧukəm “Bag”, “Bag”, “Bag!”,

(b) takar ŋa ʧukəm ŋo kaːpʰa ŋoma. “Takar I Can’t Find my Bag”

(c) hogo ŋa ʧukəm ? “Where is My Bag?”

Here the target words are Monosyllabic in this example Shown in Figures 4 and 5. Similarly disyllabic and Multisyllabic sound words in Nyishi are recorded in the same manner in this Read Mode recording of Speech. Storytelling and Number-counting sessions were also included in read mode to capture the oral traditions and narrative styles specific to the Nyishi community. Most of the speech recording was captured from diverse locations and Nyishi community social Settings, to understand how the Nyishi language is used and adapted in different environments.

To capture the clean audio that can suit to speech technology model or phonetic research etc? We opted for different devices which were used in different modes to record the speech shown in Table 1. Key standard setting was made for speech recording, Audio quality was set at a high sampling rate of 44.1 Hz with 16-bit depth which was captured at close-talk microphones mostly in either studios or quiet rooms where ambit noise is less than 50 dB SPL.

In the Spontaneous speech mode, a total duration of 10 hours was recorded. It involved 24 male and 10 female speakers from the districts of Kradaadi, Kurung Kumey, Lower Subansiri, Papumpare, and parts of Kamle Fig.4. In the Fluent speech mode, a total duration of 5 hours was recorded. It involved 14 male and 8 female speakers from the same districts. In the Read mode, a total duration of 6 hours was recorded. 6 male speakers and 4 female speakers were involved as shown in Table 2.

Table 2. Speakers Data Collection Mode

Sl.No	Mode of Speech	No. of Speakers	No. of males	No. of Females	Duration
1	Spontaneous speech	34	24	10	10 Hrs
2	Fluent speech	22	14	8	5 Hrs
3	Read Mode	10	6	4	6 Hrs

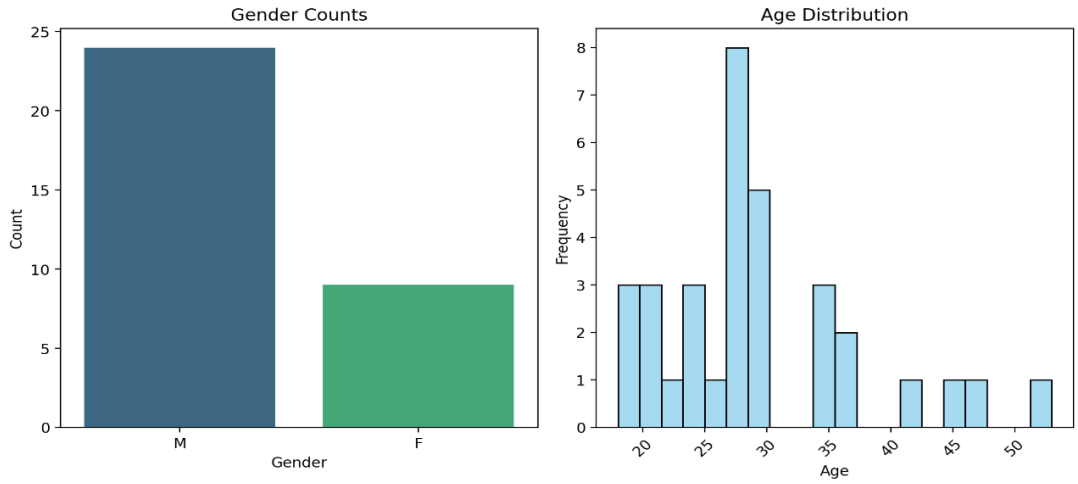

Fig 2. Speakers' Age and Gender Distribution

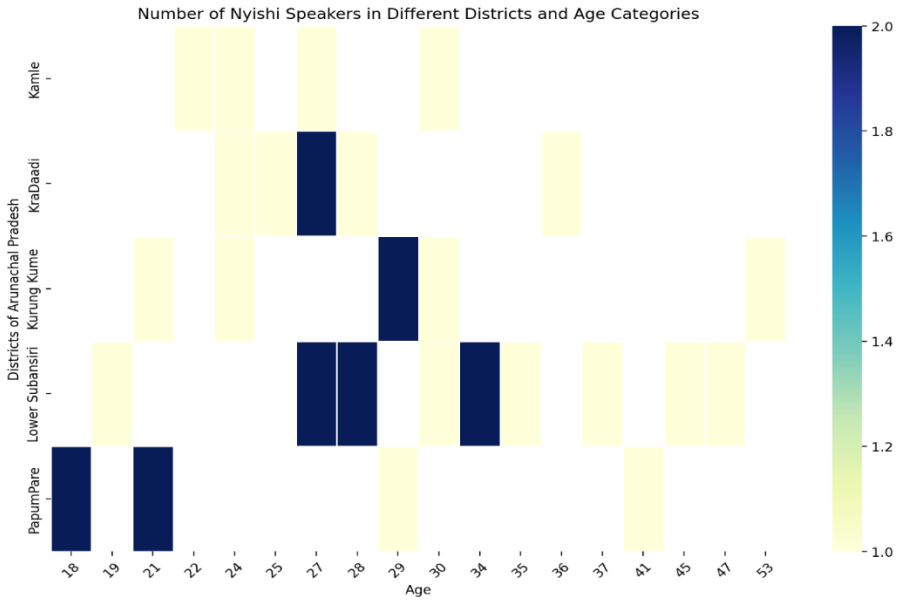

Fig 3. Speakers of Nyishi Language by Districts and Ages.

The above Fig 3. Provides a detailed view of the age distribution of Nyishi speakers across different districts of Arunachal Pradesh. It highlights the diversity of the speakers in terms of age and geographical distribution. The Nyishi speakers span a wide range of ages, from young adults to older individuals. This is evident from the vertical spread of points in each district. This is indicative of demographic trends within these districts. The density of the box gives us an idea about the number of Nyishi speakers in each district. Districts with more boxes have a higher number of speakers. In Fig 4. we see there are 5 male speakers from Kradaadi District aged around 28 years and 1 female speaker of age 36 years and from Papumpare District we have 3 male Nyishi speakers of age range from 21 to 40 years and 3 females of age range 18 to 29 years. In Kurung kume we have 6 Nyishi speakers age range from 21 to 53 years as shown in Fig 4. From the lower Subansiri District, we

have 6 males aged from 27 to 48 years and 6 females aged 19 to 45 years. Similarly, from Kamle District, only 4 male speakers are there with an age range from 24 to 30 years. The distribution of ages varies across districts. Some districts have a more concentrated age range, while others have a broader age range.

4. ANALYSE NYISHI SPEECH CORPUS

Nyishi displays rich phonological and syntactic features as a Tibeto-Burman language yet risks endangerment without modern documentation and revitalization efforts. Constructing a robust Nyishi speech corpus holds profound value for enabling language preservation initiatives, acoustic-phonetic research, and speech technology development by providing foundational speech recognition and synthesis training data.

4.1. Phonetic Composition

Nyishi, situated within the greater Sino-Tibetan language family, exhibits a phonetic richness that reflects the linguistic and cultural diversity of the Nyishi people. Major element of the phonology of the Nyishi language is its consonantal system, which includes a variety of sounds such as voiceless and voiced stops, nasals, fricatives, affricates, liquids, and glides.

Table 3. Nyishi Phonetic Representation in IPA and Equivalence ILSL12 Label.

Phonetic Symbol (IPA label)	Equivalence ILSL12 Label	Nyishi words	Phonetic Symbol (IPA label)	Equivalence ILSL12 Label	Nyishi words
a	**a**	***atʃʰ*** *"Brother"*	***tʃʰ***	**ch**	***Choqyum*** "beard"
aː	**aa**	***aːt*** *"Sister"*	***ʤ***	**j**	***Jichonam*** "advance"
e	**e**	***Elyap*** *"door"*	***ʤʰ***	**jh**	***Jhaːma*** "bore"
eː	**ee**	***eːh*** "Bamboo shoot"	***t***	**t**	***Tab*** "bug"
i	**i**	***iǵin*** "Bamboo bag"	***tʰ***	**th**	***Thaqt*** "breadth"
iː	**ii**	**i**:**rin** "anus"	***d̪***	**d**	***Dagnam*** "standing"
			d̪ʰ	**dh**	***Dhonam*** *"sitting"*
o	**o**	***Ogh*** "hot"	***n***	**n**	***Nur*** "smell"
oː	**oo**	***Oːk*** " leaf"	***p***	**p**	***Paqnam*** "appreciation"
			pʰ	**ph**	***Phehaq*** "soybean"
ə	**ə,ea**	***əǵːiqʔ*** " quiver"	***b***	**b**	***Bur*** "crocodile"
			bʰ	**bh**	***Bhayam*** "tenda"
əː	**əə**	***əːr*** "fasting"	***m***	**m**	***Mui*** "ashes"
u	**U,ʊ**	***uŋ*** " hole"	***y***	**y**	***Yamdiq*** "chilly"
uː	**uu**	***uːn*** "wound"	***r***	**r**	***rəːyi*** "core"
ɨ	**eu**	***ɨnam*** "walk"	***l***	**l**	***Labang*** "knee"
			lʰ	**lh**	***Lhagpo*** "arm"
ɨː	**ɨi,eu**:	***ɨː*** "grass"	***s***	**s**	***Səcal*** "wolf"
io	**io**	***taːbio*** "jhum hut"	***h***	**h**	***Husse*** "yam"
ui	**ui**	***ui*** "blood"	***q***	**q**	***Poroq*** "chicken"
ai	**ai**	***taiyaq*** "millet"	***x***	**x**	***Xinam*** "admire"
			Xʰ	**xh**	***Xhiqkato*** "count"
oi	**oi**	***ʃoi*** "tonight"	***z***	**z**	***Zarg*** "1 Kg"
iʊ	**iʊ,iu**	***Koriu*** "Hoe"			
ao	**ao**	***raonam*** "guard"			
əi	**əi**	***məiɬanam*** "conquer"			
əo	**əo**	***məonam*** "advance work"			
k	**k**	***Ked*** "land"			
kʰ	**kh**	***Khopaq*** "banana"			
g	**g**	***gam*** "language"			
gʰ	**gh**	***ghamro*** "loud"			
ɲ	**ny**	***ɲem*** "girl"			
ŋ	**ng**	***ŋui*** "fish"			
tʃ	**c**	***Cuk*** "bag"			

Some voiceless stops convey meaning in the nyishi language. For example, the sound k [k], kh [kʰ]: "K," the voiceless velar stop in Nyishi meaning land, "Ked" is the example. Additionally, "Khopaq" or "Kh," the voiceless aspirated velar stop indicating banana, is another notable example. This aspiration is the And t [t], th [tʰ]: Through example, "Tab" symbolizes "bug", while "Thaqt" stands for "the breadth". These variations demonstrate that Nyishi's phonetics cannot be overlooked. Again for sound p [p], ph [pʰ]: In "Paqnam," the sound of 'p' means appreciation while 'ph' in "Phehaq" signifies soybean bean. This reflects the subtle phonetic differences that make up Nyishi language. Voiced Stops: 'd' and 'dh'. The voiced dental stop 'd̪' is evident in "dagnam", which mean standing and the voiced aspirated dental stop 'dh' characterizes 'donam', meaning sitting. This implies that voice and aspiration have significance in Ny Nasals: n[n], m[m], ŋ[ŋ], ɲ[ɲ]: Voiced dental nasal n features in "Nur", the word for "smell". Voiced bilabial nase m appears in "Mui" which means 'ashes'. Voiced velar The Nyishi word "nym" for girl, for example, is made using 'ny' which represents a voiced palatal nasal. Fricatives: s[s], z[z], ʃ[ʃ]. For example, 'Səcal' means 'wolf'. 'Zarg' means 1 Kg. 'ʃoi' conveys 'tonight'. These fricatives make Nyishi more diverse phonetically.

Affricates: tʃ [tʃ] and j [ʤ]: The voiceless postalveolar affricate 'tʃ' appears in Choqyum or "beard." The voiced postalveolar affricate 'ʤ' in Jichonam or "adv w [w], j [j]: 'W' in "Koriu" (hoe), and 'J' in "jichonam" (advance) represent the voiced bilabial approximant and voiced palatal approximant respectively, contributing to the fluid pronunciation of Nyishi.. Liquids: l [l], lh [lʰ], r [r]: Labang means "knee," and it has the voiced alveolar lateral approximant 'l'; lhagpo, that implies "arm," with the voiceless alveolar lateral. Nyishi's speech has some liquids, called "r" as voiced alveolar tap or flap which make it melodious. Glide y [y]: The voiced palatal glide 'y' is evident in "Yamdiq," conveying "chilly." This glide adds a dynamic element to the Nyishi phonetic repertoire. These distinctions play a role, in Nyishi's pronunciation. And for Nyishi vowels we found that there are seven short vowels / a,i,o,e,u,ə,ɨ / and long seven vowels / əː,ɨː,aː,eː,iː,uː,oː / and some diphthongs were found [iʊ],[ao],[əi],[əo],[ai],[oi], [ui],[io] as shown on Table 3. The consonants are divided into different categories: voiceless, voiced, and aspirated. The voiceless consonants are /p/, /tʃ/, /k/, and /kʰ/. The voiced consonants are /b/, /dʒ/, /g/, /m/, /n/, /ɲ/, /ŋ/, and /l/. The aspirated consonants are /pʰ/, /tʃʰ/, /kʰ/, and /ɑːkʰ/.

The Nyishi language, also known as Nyising, it is one of the major languages of the Tani subgroup within the Tibeto-Burman language family [25]. The phonetic composition and phonological structure of Nyishi have been examined by several linguists in recent decades. [5] provides an updated analysis of Nyishi tones, identifying five contrastive tones - high level, mid-level, low level, high falling, and low falling. However, Nyishi is still considered a low-resource language due to the limited available linguistic resources. Nyishi faces significant challenges due to a lack of digital resources, such as existing speech datasets or transcribed text data. Without sufficient resources, studying and documenting the phonetic and phonological patterns of low-resource languages like Nyishi can be difficult. This study aims to help address the need for more linguistic resources by creating a speech corpus for the Nyishi language. Developing annotated Nyishi speech datasets will provide data to enable more accurate phonetic and phonological analysis of this low-resource Tibeto-Burman language. Nyishi Language sounds are unique and different from other languages. Nyishi Language doesn't have any designated script, some of the sounds can be represented in devnagri style, but again few Nyishi phoneme sounds cannot be represented in devnagri representation. Therefore most of the Nyishi sounds are being transcribed in IPA Based Transcription. During the recording of Nyishi's speech in different modes, Number counting sessions were also included in read mode to capture the oral traditions and narrative styles specific to the Nyishi counting system in the Nyishi community shown in Table 4. With IPA IPA-based transcription and some devnagri representation.

Table 4. Nyishi Numerals 1to100 in IPA and Devanagari representations.

N-C	Nyishi / Devnagri	N-C	Nyishi / Devnagri	N-C	Nyishi / Devnagri	N-C	Nyishi / Devnagri	N-C	Nyishi / Devnagri	N-C	Nyishi / Devnagri	N-C	Nyishi / Devnagri	N-C	Nyishi / Devnagri	N-C	Nyishi / Devnagri
0	saːr शून्य																

7	6	5	4	3	2	1
kən	aːkʰə	aːŋ	əpy	oːum	əɲI	aːkəin
कन	आख	आङ	-	ओउम	-	आकीन
17	16	15	14	13	12	11
əriːjəŋleːk- kən	ə riːjəŋleːk-kʰi	əriːjəŋleːk-ŋo	əriːjəŋleːk- py	əriːjəŋleːk- um	əriːjəŋleːk-ɲI	əriːjəŋleːk- kəin
--	--	--	--	--	--	--
27	26	25	24	23	22	21
caːməŋleːk- kən	caːməŋleːk- kʰi	caːməŋleːk- ŋo	caːməŋleːk- py	caːməŋleːk- um	caːməŋleːk- ɲI	caːməŋleːk- kəin
चामनयले कन	चामनयले आख	चामनयले आङ	चामनयले--	चामनयले ओउम	चामनयले--	चामनयले कीन
37	36	35	34	33	32	31
caːmumleːk- kən	caːmumleːk- kʰi	caːmumleːk- ŋo	caːmumleːk- py	caːmumleːk- um	caːmumleːk- ɲI	caːmumleːk- kəin
चामुमलेक़ कन	चामुमलेक़ आख	चामुमलेक़ आङ	चामुमलेक़--	चामुमलेक़ ओउम	चामुमलेक़--	चामुमलेक़ कीन
47	46	45	44	43	42	41
caːməpəjə leːk- kən	caːməpəjə leːk- kʰi	caːməpəjə leːk- ŋo	caːməpəjə leːk- py	caːməpəjə leːk- um	caːməpəjə leːk- ɲI	caːməpəjə leːk- kəin
चामपयलेक़ कन	चामपयलेक़ आख	चामपयलेक़ आङ	चामपयलेक़--	चामपयलेक़ ओउम	चामपयलेक़--	चामपयलेक़ कीन
57	56	55	54	53	52	51
caːmŋo leːk- kən	caːmŋo leːk- kʰi	caːmŋo leːk- ŋo	caːmŋo leːk- py	caːmŋo leːk- um	caːmŋo leːk- ɲI	caːmŋo leːk- kəin
चामनगोलेक़ कन	चामनगोलेक़ आख	चामनगोलेक़ आङ	चामनगोलेक़--	चामनगोलेक़ ओउम	चामनगोलेक़ --	चामनगोलेक़ कीन
67	66	65	64	63	62	61
caːməkʰə leːk- kən	caːməkʰə leːk- kʰi	caːməkʰə leːk- ŋo	caːməkʰə leːk- py	caːməkʰə leːk- um	caːməkʰə leːk- ɲI	caːməkʰə leːk- kəin
चामखलेक़ कन	चामखलेक़ आख	चामखलेक़ आङ	चामखलेक़--	चामखलेक़ ओउम	चामखलेक़--	चामखलेक़ कीन
77	76	75	74	73	72	71
caːməkən leːk- kən	caːməkən leːk- kʰi	caːməkən leːk- ŋo	caːməkən leːk- py	caːməkən leːk- um	caːməkən leːk- ɲI	caːməkən leːk- kəin
चामकनलेक़ कन	चामकनलेक़ आख	चामकनलेक़ आङ	चामकनलेक़--	चामकनलेक़ ओउम	चामकनलेक़--	चामकनलेक़ कीन
87	86	85	84	83	82	81
caːməkən leːk- kən	caːməkən leːk- kʰi	caːməkən leːk- ŋo	caːməkən leːk- py	caːməkən leːk- um	caːməkən leːk- ɲI	caːməpəiːn leːk- kəin
चामपईनलेक़ कन	चामपईनलेक़ आख	चामपईनलेक़ आङ	चामपईनलेक़--	चामपईनलेक़ ओउम	चामपईनलेक़--	चामपईनलेक़ कीन
97	96	95	94	93	92	91
caːməkiːjaː leːk- kən	caːməkiːjaː leːk- kʰi	caːməkiːjaː leːk- ŋo	caːməkiːjaː leːk- py	caːməkiːjaː leːk- um	caːməkiːjaː leːk- ɲI	caːməkiːjaː leːk- kəin
चामकीयालेक़ कन	चामकीयालेक़ आख	चामकीयालेक़ आङ	चामकीयालेक़--	चामकीयालेक़ ओउम	चामकीयालेक़--	चामकीयालेक़ कीन

10	9	8
ǝri:jǝŋ	kija:	pi:n
--	किया	पीन
20	19	18
ca:mǝɲ	ǝri:jǝŋle:k- kija:	ǝri:jǝŋle:k- pi:n
चामनय	--	--
30	29	28
ca:mum	ca:mǝɲle:k- kija:	ca:mǝɲle:k- pi:n
चामुम	चामनयले किया	चामनयले पीन
40	39	38
ca:mǝpǝjǝ	ca:mumle:k- kija:	ca:mumle:k- pi:n
चामपय	चामुमलेक किया	चामुमलेक पीन
50	49	48
ca:mɲo	ca:mǝpǝjǝ le:k- kija:	ca:mǝpǝjǝ le:k- pi:n
चामनगो	चामपयलेक किया	चामपयलेक पीन
60	59	58
ca:mǝkʰǝ	ca:mɲo le:k- kija:	ca:mɲo le:k- pi:n
चामख	चामनगोलेक किया	चामनगोलेक पीन
70	69	68
ca:mǝkǝn	ca:mǝkʰǝ le:k- kija:	ca:mǝkʰǝ le:k- pi:n
चामकन	चामखलेक किया	चामखलेक पीन
80	79	78
ca:mǝpǝi:n	ca:mǝkǝn le:k- kija:	ca:mǝkǝn le:k- pi:n
चामपईन	चामकनलेक किया	चामकनलेक पीन
90	89	88
ca:mǝki:ja:	ca:mǝkǝn le:k- kija:	ca:mǝkǝn le:k- pi:n
चामकीया	चामपईनलेक किया	चामपईनलेक पीन
100	99	98
ca:mǝɹji:ŋ or lǝŋ or liŋ	ca:mǝki:ja: le:k- kija:	ca:mǝki:ja: le:k- pi:n
चामऋयीङ	चामकीयालेक किया	चामकीयालेक पीन

Note: -N-C stands for Number counting.--- cannot represent on Devnagri sound system.

The Nyishi number system presents a fascinating array of phonetic features, encapsulating the linguistic richness of the Tibeto-Burman language spoken primarily in Arunachal Pradesh, India. The Nyishi number system features unique phonetic compositions that are crucial for understanding and pronouncing the numbers correctly as shown in Table 4., There is a variety of vowel sounds, including [ɑː], [ə], [oː], [ɨ], and [iː]. These vowels contribute to the distinctiveness of the language. For example, the number '1' is pronounced as [ɑːkəin], where [ɑː] represents the long open front unrounded vowel. Nasalization in [ɑːŋ] involves the velar nasal [ŋ] as shown in [ɑːŋ] (5). Aspiration [kʰə] requires a burst of air during articulation in [ɑːkʰə] (6). The nasalized vowel [iː] and the unaspirated voiceless bilabial plosive [p] create a distinct sound in [piːn] (8). The complex cluster [jəŋleːk-k] involves multiple consonants in sequence in [əriːjəŋleːk-kəin] (11). In [cɑːməɲleːk-kən] (27), the presence of clusters ([ɲleːk-k]) showcases Nyishi's intricate phonetic structures, and similarly in [əriːjəŋleːk-py] (14) the cluster [jəŋleːk-p] involves a transition from [ŋ] to [p]. Nasalization in [ŋo] adds a distinctive quality as shown in [əriːjəŋleːk-ŋo] (15). The sound [ɲ] occurs in [cɑːməɲ](20), contributing to the language's phonetic richness. The rising tone in [cɑːməɲleːk-kəin] emphasizes the pitch change in Nyishi shown on [cɑːməɲleːk-kəin] (21) in Table 4. Unique pronunciations ([ɲI]) emphasize Nyishi's distinct phonetic features in [cɑːməɲleːk-ɲI] (22). And in [cɑːməpəjə leːk-um] (43) Smooth transitions in [pəjə leːk-um] involve both nasalized and non-nasalized vowels. The long vowel sound /ɑː/ is present in numbers like "ɑːkə" (6), while the schwa sound /ə/ is found in "kən" (7). Nasal sounds like /n/ and /ɲ/ appear in "kən" (7) and "əɲI" (2). There is variation in the consonant sounds, including aspirated consonants like /kʰ/ in "ɑːkʰə" (6) and nasal sounds like /ɲ/ in "əɲI" (2). Aspirated consonants are present, as in "ɑːkʰə" (7), where /kʰ/ indicates the aspirated form of /k/. Beyond basic cardinal numbers, compound words are formed to represent numbers beyond ten. For instance, "ɑːkəin" (10) is a compound of "ɑːkə" (6) and "kən" (1). Some numbers are single syllables (e.g., "kən"), while others are compound words with multiple syllables (e.g., "əriːjəŋleːk- kəin"). Affixes are added to the base numbers to represent multiples of ten. "əriːjəŋleːk-kəin" (11) adds "-leːk" to "ɑːkəin" (10). Palatalization occurs in "cɑːməɲ" (20), where /cɑːm/ features a palatalized consonant /c/. The numbers generally follow a pattern of one or two syllables, contributing to the rhythmic and melodic quality of Nyishi counting. Examples: "kən" (1), "ɑːkəin" (10).Some numbers, especially those beyond ten, have complex pronunciation due to combinations of sounds and affixes. Examples: "əriːjəŋleːk-kəin" (11), "əriːjəŋcɑːm" (21).Certain numbers have unique terms that don't follow a predictable pattern. Memorizing these terms is essential, such as "cɑːməɲ" for "20." Some numbers have complex pronunciation due to the combination of sounds and

affixes, making them distinct from simple cardinal numbers.

4.2. Transcription

The methods used in transcribing Nyishi speech involve a combination of listening to and analyzing audio recordings, understanding the grammar and vocabulary rules of the language, and creating a written representation of the spoken words and sentences. - Specialized software and tools were used to assist in the transcription process. Praat software was used to analyze the acoustic properties of the Nyishi speech signals and aid in accurate transcription. Three layers of transcription were made mostly in IPA Based Transcription as shown in Fig. 4 and 5.

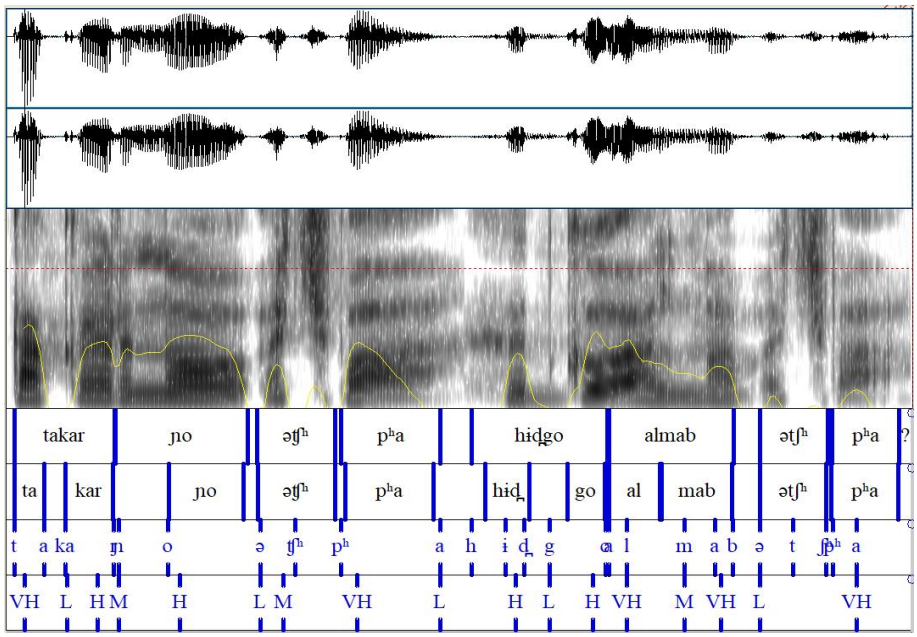

Fig 4. Nyishi Sentences "*takar ɲo əʧʰ pʰa hiɖgo almab əʧʰ pʰa* " in three layer Transcription.

In Nyishi sentence Fig 4. "*takar ɲo əʧʰ pʰa hiɖgo almab əʧʰ pʰa* " which means *Takar are you sick, how bad is it ?*And Similarly in Sentence two Fig 5. *"ʧuk, ʧuk, ʧukəm takar ŋa ʧukə ŋo ka:pʰa ŋoma, hogo ŋa ʧukəm ? ""Bag", "Bag", "Bag!", "Takar I Can't Find my Bag". "Where is My Bag?"* The first level of Transcription is phonetic transcription (representation of speech sounds in visual form in praat and using IPA symbols or letters to represent individual sounds.), Second is syllable level transcription (separating words into syllables, which helps in understanding the rhythm and structure of speech.), and Third is the Phoneme Pitch level transcription (It involves estimating the pitch of the speech by observing the pitch contour, which is the pattern of pitch changes over time shown in Fig. 4 and 5) to capture the nuances of the Nyishi language accurately. Pitch frequency is different for different people's age groups. By considering a particular age group person's frequency variation depending on his speech production vocal tract system and environment, pitch contour is estimated in different levels, levels are High (H), Low (L), Very high (VH) and Mid (M), and these levels are fixed by visual observation of frequencies in pitch contour. Break marking is done in Praat. It involves identifying pauses in the speech signal by examining regions of voiced and unvoiced sounds, as well as listening to the sound.

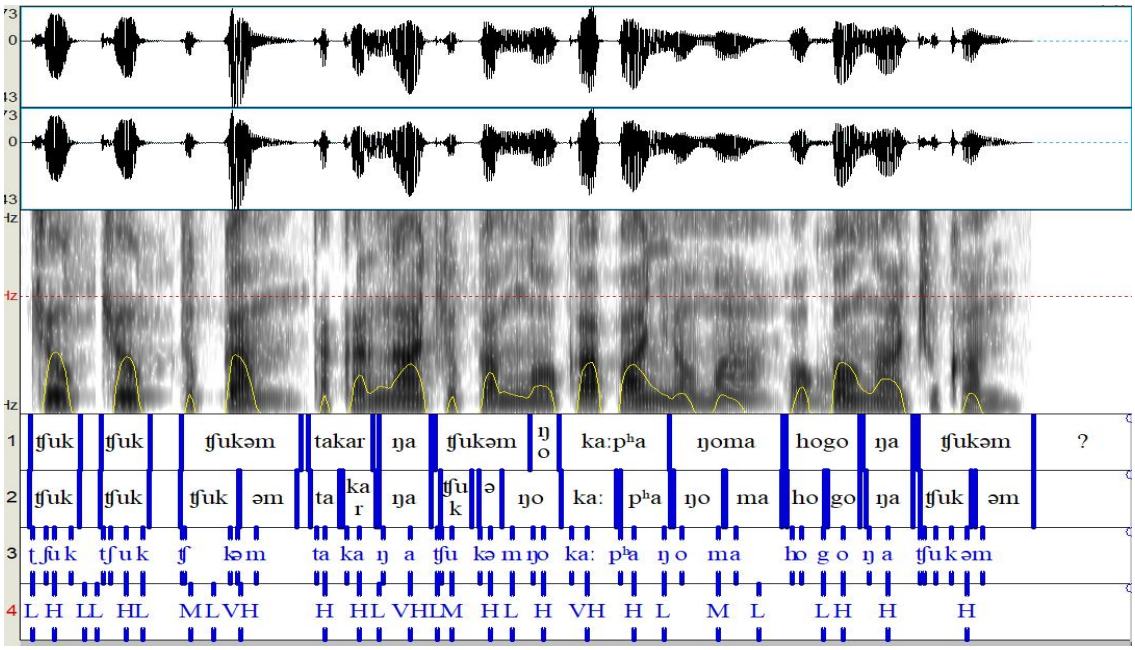

Fig 5. Nyishi Sentences "*ʧuk, ʧuk, ʧukəm takar ŋa ʧukə ŋo ka:pʰa ŋoma, hogo ŋa ʧukəm* ? "in three layer Transcription.

4.3 Statistical Analysis of Nyishi Speech Corpora.

In examining the frequencies of Nyishi vowels evident in the corpus, the vowel /a/ appears most commonly with 22,394 total instances covering 34.1% in the corpus shown in Figures 6 and 7. This predominance of /a/ likely relates to its central role as the default or schwa-like vowel in the Nyishi phonological system. The next most frequent vowels are /i/ at 7,679 with 11.7% instances and /o/ at 9,142 with 13.9% instances. The front close vowel /i/ functions prominently in Nyishi words, often appearing in initial and medial syllables, while /o/ marks many final syllables functioning as a grammatical marker. Less common are the vowels /e/ at 2,440 in 3.7% instances in the corpus, /u/ at 5,308 with 8.1% instances, and /ə/ at 5,598 in 8.5% instances. The mid-front vowel /e/ enables certain diphthongs and tripthongs, while /u/ commonly marks grammatical categories. The schwa vowel /ə/ typically appears in initial or medial positions. Even more infrequent vowels attested are the central vowels /ɨ/ at 6,210 with 9.5% instances and its longer counterpart /ɨː/ at 1,262 instances. These two central vowels distinguish meanings in minimal pairs and occur in initial, medial, and final syllables. Rounding out the Nyishi vowel inventory evident are the lengthened vowels /əː/ at 551 instances, /aː/ at 1,749 instances, /eː/ at 313 instances, /iː/ at 280 instances, /uː/ at 446 instances and /oː/ at 2,291 instances. These elongated vowels carry phonemic contrasts, increase syllable weight, and indicate certain grammatical categories.

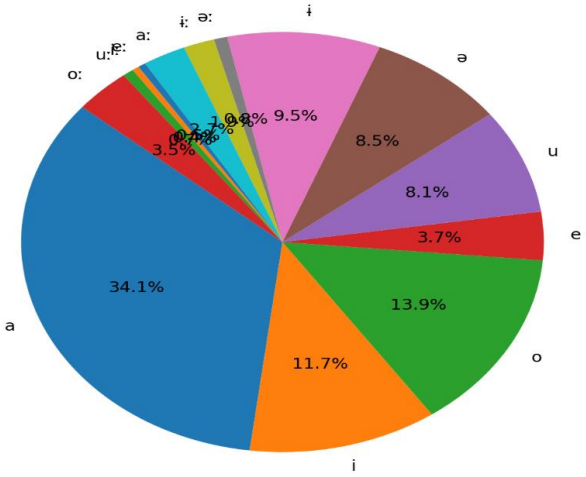

Fig 6. Proportion of Nyishi Vowels in Corpus

Each slice of the pie in Fig 6. represents a vowel or a long vowel, and the size of the slice corresponds to the proportion of the vowel in the column. The labels show the exact percentage of each vowel.

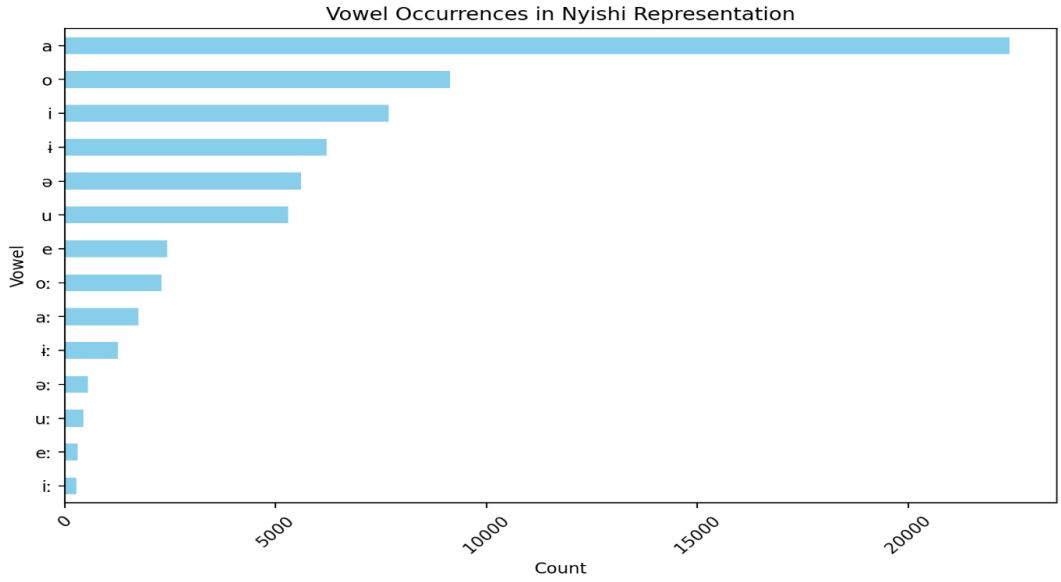

Fig 7. Nyishi vowel frequencies count in the Corpus

This Fig 6 & 7 visualization provides a clear view of the relative frequencies of the vowels and long vowels in the 'Nyishi Representation' column. For instance, we can see that 'a' is the most common vowel, while 'əː' and 'ɨː' are among the least common, Similarly in Fig 7 the bar graph shows the total counts of each Nyishi vowel in the corpus. The x-axis represents the different vowels, and the y-axis represents the count of each vowel. From the graph, we can see that the vowel 'a' has the highest count, followed by 'i'. The vowel ' iː ' has the lowest count.

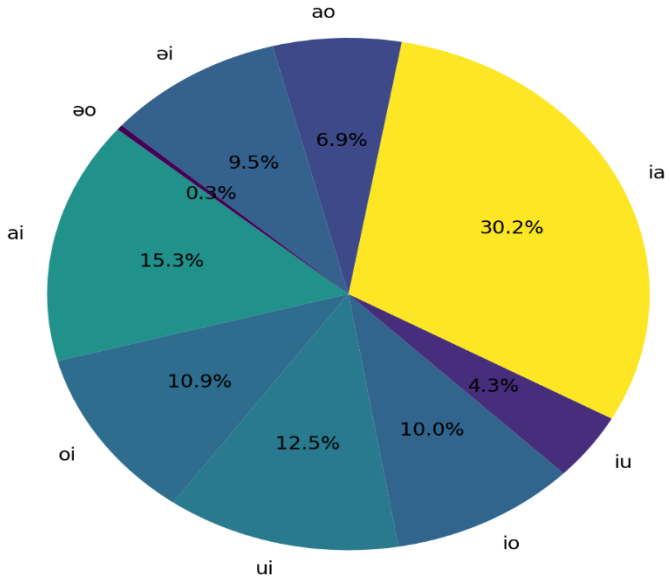

Fig 8. Proportions of Nyishi Diphthongs in Corpus

Nyishi speech corpus displays productive use of diphthongs in the language. In Fig 8. Each slice of the pie represents a diphthong, and the size of the slice corresponds to the proportion of the diphthong in the column. The labels show the exact percentage of each diphthong. The most frequently appearing diphthong is /ai/ at 187 total with 15.5% instances, formed by the combination of /a/ and /i/. Other more moderately evidenced combos are /oi/ at 133 with 10.9% instances, blending /o/ and /i/, along with /ui/ at 153 with 12.5% cases joining /u/ and /i/. Examples of less common Nyishi diphthongs captured include /io/ at 122 with 10% instances, mixing /i/ and /o/, and /iu/ at 52 covering 4.3% examples combining /i/ and /u/. Rounding out the inventory are the triphthong-like sequences of /ia/ at 369 instances covering highest 30.2%, /ao/ at 84 examples covering 6.9%, /əi/ at 116 cases in 9.5%, and /əo/ at 4 lowest 0.3% total attestations in the corpus. These patterns reveal a predominant reliance on front vowels /i/ and /e/ when forming Nyishi diphthongs and complex

vocalic sequences. This Fig 7. visualization provides a clear view of the relative frequencies of the diphthongs in the 'Nyishi Representation' column. For instance, we can see that 'ia' is the most common diphthong, while 'ɔo' is the least common.

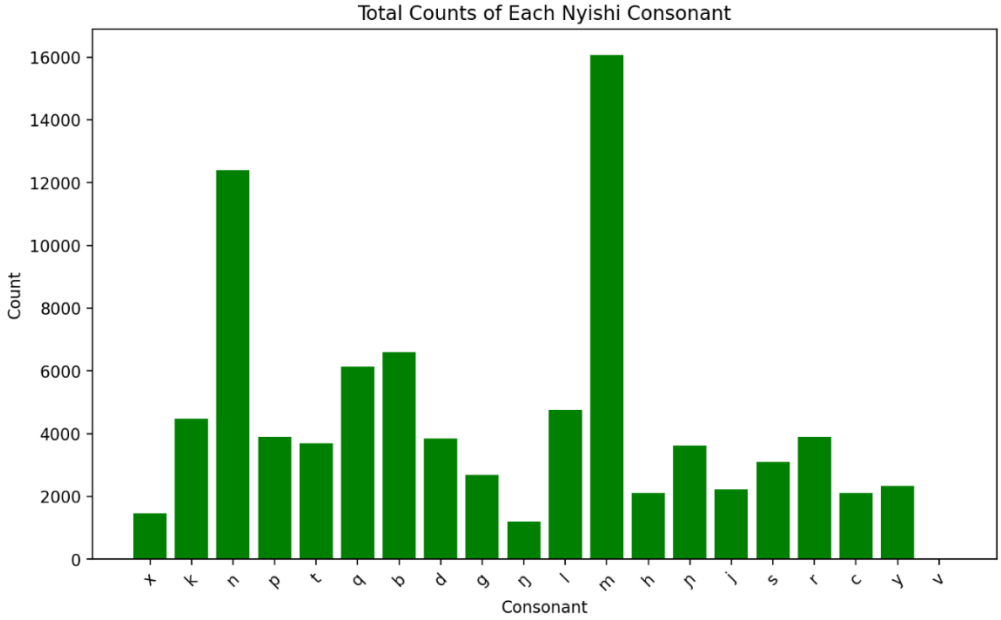

Fig 9. Nyishi Consonant frequencies count in Corpus

Similarly, the Fig 9. The bar chart shows the proportion of each Nyishi consonant in the corpus. Each slice of the chart represents a different consonant, and the size of the slice corresponds to the number of the total counts that each consonant represents. From the chart, we can see that the consonant 'm' makes up the largest proportion of the total counts, followed by 'n'. The consonant 'v' makes up the smallest proportion.

5. Conclusions

This paper has provided an overview of the ongoing efforts to develop a Nyishi speech corpus to support research and technology development for this low-resource Nyishi Language one of the Tibeto-Burman languages. Constructing the corpus involved thoughtful data collection methodologies like lab-based recording sessions, district-wise visit recording as well as some crowd-sourced speech samples to ensure diversity of speakers and dialects. 34 native Nyishi speakers across gender, age ranges, and regional backgrounds contributed recordings in various speech modes including spontaneous conversations, fluent speech, read speech, and storytelling narratives. The resulting corpus showcases the rich phonetic composition of Nyishi, which features seven vowels, seven corresponding long vowels, a range of consonants, and productive diphthongs and triphthongs. Statistical analysis of phoneme distributions reveals the central role of /a/ as the most common vowel at 34.1% of instances and the peripheral status of vowels like /əː/ and /iː/. Meticulous transcription was carried out in International Phonetic Alphabet symbols with markings for tone, pitch, syllabification, and pacing. The paper outlines the specific file naming conventions developed to enable computational parsing of speaker demographics and recording conditions encoded into the file names themselves. This corpus will serve as a valuable resource for both computational and human linguistic research. Applications span from acoustic-phonetic analysis, speech recognition, and synthesis model development to dialectology studies and investigations of Nyishi tonality, lexicon, morphology, and phrasing patterns. While this paper has summarized progress so far in constructing the early Nyishi speech corpus, there remain opportunities for expansion and enhancement. Priorities include augmenting corpus size through additional recording sessions, potentially involving new dialects, speech genres, and phrase types. Enriching speaker demographic diversity in terms of age, gender, and sub-region would also bolster corpus value.

6. ACKNOWLEDGEMENT

The Author would like to extend our sincere appreciation to all the native Nyishi speakers who volunteered their time and voices to make this speech corpus a reality. We are grateful to the 34 participants hailing from the Kradaadi, Kurung Kumey, Lower Subansiri, Papumpare, and Kamle districts of Arunachal Pradesh. Their enthusiasm to help document and revitalize their ancestral language demonstrates the resilient Nyishi culture. This corpus would not have been achievable without the goodwill of members of the Nyishi community. We are thankful for their efforts to

preserve Nyishi's heritage for generations to come through these recordings.

Conflict of Interest

The authors declare that they have no conflicts of interest.

References

[1] Driem, G. v. "Lost in the sands of time somewhere north of the bay of Bengal. Himalayan Languages and Linguistics", 11-38(2011) . https://doi.org/10.1163/ej.9789004194489.i-322.10.

[2] Thurgood, Graham, and Randy J. LaPolla, eds. The Sino-Tibetan languages. Taylor & Francis, 2016.

[3] Driem, G. v."The diversity of the tibeto-burman language family and the linguistic ancestry of chinese. Bulletin of Chinese "Linguistics,1(2)(2007),211270.https://doi.org/10.1163/2405478x90000023.

[4] Post, Mark W. "Tones in Northeast Indian languages, with a focus on Tani: A fieldworker's guide." Language and culture in Northeast India and beyond: In honour of Robbins Burling (2015): 182-210.

[5] Gauthier, Elodie, Laurent Besacier, and Sylvie Voisin. "Automatic speech recognition for African languages with vowel length contrast." Procedia Computer Science 81 (2016): 136-143.

[6] Godfrey, John J., Edward C. Holliman, and Jane McDaniel. "SWITCHBOARD: Telephone speech corpus for research and development." Acoustics, speech, and signal processing, ieee international conference on. Vol. 1. IEEE Computer Society, 1992.

[7] Singh, Amitoj, et al. "ASRoIL: a comprehensive survey for automatic speech recognition of Indian languages." Artificial Intelligence Review 53 (2020): 3673-3704.

[8] Jia, Ye, et al. "CVSS corpus and massively multilingual speech-to-speech translation." arXiv preprint arXiv:2201.03713 (2022).

[9] Poria, Soujanya, Erik Cambria, and Alexander Gelbukh. "Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis." Proceedings of the 2015 conference on empirical methods in natural language processing. 2015.

[10] Panayotov, Vassil, et al. "Librispeech: an asr corpus based on public domain audio books." 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2015.

[11] Corpus, VoxForge Russian Speech. "Retrieved January 15, 2010, from VoxForge: http://www. dev. voxforge. org/projects/Russian/browser/ Trunk/AcousticModels." (2007).

[12] Lane, Ian, et al. "Tools for collecting speech corpora via Mechanical-Turk." Proceedings of the NAACL HLT 2010 workshop on creating speech and language data with amazon's mechanical turk. 2010.

[13] Novotney, Scott, and Chris Callison-Burch. "Cheap, fast and good enough: Automatic speech recognition with non-expert transcription." Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. 2010.

[14] Levine, Sergey, et al. "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection." The International journal of robotics research 37.4-5 (2018): 421-436.

[15] Jia, Ye, et al. "Direct speech-to-speech translation with a sequence-to-sequence model." arXiv preprint arXiv:1904.06037 (2019).

[16] Dondrup, R. A handbook of the Nyishi language. Itanagar, India: Directorate of Research, Govt. of Arunachal Pradesh (1988)..

[17] Abraham, B. Word tones in Nyishi. Indian Linguistics(1985)., 46, 19-30.

[18] Robinson, Tony, et al. "WSJCAMO: a British English speech corpus for large vocabulary continuous speech recognition." 1995 International Conference on Acoustics, Speech, and Signal Processing. Vol. 1. IEEE, 1995.

[19] Deka, Barsha, et al. "Speech corpora of under resourced languages of north-east india." 2018 Oriental COCOSDA-International Conference on Speech Database and Assessments. IEEE, 2018.

[20] Kumar, Gopendra. Geology of Arunachal pradesh. GSI, 2013.

[21] Feng, R., & Guo, Q. Second Language Speech Fluency: What Is in the Picture and WhatIsMissing (2022, February 28).