Using forced alignment for sociophonetic research on a minority language

Danielle Barth, James Grama, Simon Gonzalez and Catherine E. Travis*

1 Introduction

The advancement of current technologies has had a substantial impact on the field of linguistics, with one major change being the scope of studies conducted. In the field of phonetics, the automatic segmentation of phonemic segments from orthographic transcriptions, for example, has allowed researchers to extract and analyze vastly more data than was feasible using manual methods. Most studies applying such forced alignment have been on majority languages, for which robust acoustic models have been established based on large speech corpora, including English (Fromont and Watson 2016, Labov et al. 2013), Spanish (Goldman and Schwab 2011), and French (Brognaux et al. 2012, Milne 2014). While forced alignment has been applied to underdocumented and minority languages (e.g., DiCanio et al. 2012, Coto-Solano and Nicholas 2018), it has yet to become to be a standard tool in language documentation. This is due in large part to the understanding that a considerable amount of data is required to develop an acoustic model on the basis of which to reliably force align data. Very often, this data scale is not available for minority languages. For example, the Carnegie Mellon University Pronouncing Dictionary (Carnegie Mellon University 1993-2014) is pre-trained on 25 hours of hand-aligned speech, specifically, the SCOTUS corpus, consisting of oral arguments in the Supreme Court of the United States across a 50-year span (Yuan and Liberman 2008). More recently, it has been shown that large amounts of transcribed, spoken data are not crucial to achieving high quality forced alignment, and that there are computational steps that can be taken to increase the viability of the force-aligned output, including with relatively small speech samples (Fromont and Watson 2016, Gonzalez et al. 2018).

In this study, we describe a process for the implementation of forced alignment to a minority language (here, Matukar Panau), utilizing the Montreal Forced Aligner (McAullife et al. 2017), and test the quality of the alignment using a post-hoc algorithm. We then use the force-aligned data to conduct the first acoustic analyses of this language, considering the social conditioning of Matukar Panau vowels and, in particular, the impact of clan as a sociolinguistic variable.

2 Matukar Panau – Language Ecology and Documentation

Matukar Panau is an Oceanic language spoken on the north coast of Papua New Guinea, near the city of Madang. It exists in long-term contact with Papuan (non-Austronesian) languages of New Guinea, and has syntactic features typical of both Oceanic languages, such as serial verb constructions and direct and indirect possession, and of Papuan languages, such as clause chaining. It is an agglutinating, non-tonal language. Documentation is ongoing, and based on our current knowledge, the phonemic inventory consists of 17 consonant and 9 vowel phonemes (five monophthongs and four diphthongs).

The language is spoken in Matukar village, with around 500 people, and a smaller hamlet, Surumarang, with around 200 people. Of these 700 people, most (~540) are under 30 years old and are unlikely to speak more than very basic Matukar Panau. Their first and dominant language is the English-based creole Tok Pisin. Approximately another 130 or so people are between 30 and 50 years old. Their first language is Matukar Panau, but many no longer speak it on a regular basis. The dominant language for most of these people is certainly Tok Pisin and they have experienced

*We would like to offer a heartfelt thanks to members of the Matukar and Surumarang communities, as well as to Living Tongues Institute for Endangered Languages, National Geographic Enduring Voices, The Firebird Foundation for Anthropological Research, the ARC Centre of Excellence for the Dynamics of Language, and the Endangered Languages Documentation Program (MDP 0382) for financial support for language documentation. We also acknowledge support from an ARC Centre of Excellence for the Dynamics of Language Transdisciplinary & Innovation Grant (TIG952018). Finally, a major thank you to Sunkulp Ananthanarayan for his help in the early stages of Matukar Panau forced alignment and having the idea to start this process.

attrition of their Matukar Panau (cf. Schmid and Köpke 2007). Around 25 people are over the age of 50, and while these speakers are also Tok Pisin bilinguals, they still speak Matukar Panau often and well. Their language choice is affected by the interlocutor, as older speakers are more likely to use Tok Pisin when speaking with younger community members. Tok Pisin is also necessarily used with people from other language communities, with whom there is a great deal of contact, as the village lies next to the busy North Coast Highway, part of the primary route from Bogia to Lae.

In addition to the Matukar-Tok Pisin bilingualism, many villagers (especially older members) speak another indigenous language. There are several exogamous marriages, and thus some people have learned the language of their spouse or parent from another village. People may speak a Papuan language like Bargam, or a closely related Oceanic language like Takia, or both. Still others speak Gedaged, English, Manam, Ngain, Pelipoai, Riwo, Waskia, Widar, Yamai, or Yoidik. By the same token, some spouses of native Matukar villagers have learned Matukar Panau to some extent. The language situation is therefore complex, with prevalent multilingualism. However, even non-speakers living in Matukar and Surumarang have a strong association between Matukar Panau and belonging, as the language is exclusive to these villages. The connection between language and identity, even for semi-speakers, seems to be generally strong in this area of Papua New Guinea which has many languages spoken in rather small geographic areas.

One of the primary axes of community orientation is the clan system. Matukar has four main clans: Bantibun, Binganen, Tamaten Honen and Tamaten Painen.[1] The Matukar clans are in two groups, each of which has a "big brother" and a "little brother". Bantibun is the big brother clan to little brother Binganen, and Tamaten Honen is the big brother to little brother Tamaten Painen. Each clan group typically resides in a separate part of the village, but mixing occurs regularly between the big and little sibling clans. There is one main clan leader for the Bantibun-Binganen clan group and another for the Tamaten clan group. There is also a relatively newer clan—Kutur—that is an offshoot of Tamaten Painen. Kutur members moved to a separate area approximately 50 years ago for reasons that are unclear. Some people say this was due to disagreement with the clan, others say families simply wanted to strike out and claim their own area. Some Kutur members consider themselves their own clan, although members of the main clans dispute the status of Kutur as a separate clan and consider them to be part of Tamaten Painen.

Gender is another important axis of community organization. Both women and men tend to have more within- than across-gender contact. There are certain decisions made and rituals carried out by only one gender, and many divisions of labor exist across gendered lines. Furthermore, certain domains of knowledge, such as coming-of-age practices, are traditionally kept by only one gender. Importantly, while the people of these villages have more contact with members of their clan, their families, and their neighborhood areas, all speakers of any clan or gender have nearly daily contact with each other because of the small geographic distances between the groups.

3 Community-based Corpus Construction for an Under-documented Language

The data used for the present analysis was collected over five field trips starting in 2010. We take a community-based transcription approach. Since 2013, transcription and translation efforts have primarily been the work of the Matukar Panau Transcription and Translation team.[2] The Transcription and Translation team transcribes the data using ELAN (Lausberg and Sloetjes 2009). The members of the team are in their 20s to early 40s, and because they are not fluent speakers of the language, they will often pair with language experts to transcribe the material and translate it into Tok Pisin. The language experts are often their relatives, and working with them, the transcribers learn more Matukar Panau. Some older members of the community are then involved in language teaching. This pairing of more and less experienced community members not only enhances the quality of transcription, but also supports language maintenance and revitalization.

[1] The hamlet of Surumarang has one clan, Maran, not included in this study.

[2] Rudof Raward, Justin Willie, Alfred Sangmei, Amos Sangmei, Jillian Forepiso, Micheal Balias, and Zebedee Kreno† and the help from consultants to edit the data. The primary consultant is Kadagoi Rawad Forepiso. Other consulting help has come from Kennedy Barui†, Agnes Darr, Simporian, Cathy Samun Williang, Taleo Kreno, Berry Barui and John Bogg.

Some transcription and translation is done by a linguist (Barth). Each transcriber has their own tendencies in spelling and word boundaries, as the orthography and spelling for the language is not standardized. Workshops have been conducted in the field to establish protocols for transcription but variation remains. Transcriptions and translations are then checked and edited by the linguist and the primary consultant, Kadagoi. Editing involves standardizing spelling, editing mistranscribed words, making some substitutions for what Kadagoi considers to be grammatical errors, substituting Matukar Panau words for Tok Pisin words, and adding English translations. These kinds of changes to the transcriptions are valued by the community as material for language maintenance. The ELAN files are versioned, so that original spellings and transcriptions without any substitutions are kept, however some of the primary versions used for further analysis, including for forced alignment, have transcriptions that are not fully verbatim. While the spelling standardization helps the forced alignment process, word replacement and syntactic changes are counter-productive. Not all texts in the present collection have gone through the full and labor-intensive editing process, so there remain multiple orthographic representations of single word forms. Despite the various spellings, and despite some mismatches between audio signal and transcriptions, we were able to obtain reliably aligned data through forced alignment to perform sociophonetic analysis.

The result of the realities of documentation is that the data is not “clean” or completely consistent. This has impacts for corpus building and data analysis. Until recently, substantial work on the language has been done only by the linguist and the local team, although comparative descriptions, word lists and other brief descriptions have been published (see Anderson, Barth and Rawad Forepiso 2015, Harrison, Anderson and Barth 2010-2012, Kaspruś 1942, Z’Graggen 1969). The Matukar Panau Corpus[3] is comprised of traditional narratives, family histories, interviews, video and picture stimuli descriptions, and procedural texts. Example (1) below provides an illustration of the data from a narrative produced during an interview, describing the traditional way to tattoo.

(1)

kinkin	main,	ai	duduru-n	ngam-ngale-ma,
tattoo	TOP	tree	needle-3SG	1PL.EXCL-get-D:HAB

ngam-sututulen-dope,	takoko	ngam-ngale-ma,	ngam-rauti-nggo.
1PL.EXCL-tap.tap-D:COND1	mark	1PL.EXCL-get-D:HAB	1PL.EXCL-rub-İ:R:İPFV

y-en-dope,	gaugau-dope,	kinkn	sa-i-pid-ago
3SG-lay-D:COND1	dry-D:COND1	tattoo	ascend-3SG-DWN-R:I:PFV

‘Tattoos, we take a needle from a tree, we tap tap, we get a mark and we rub it. It is there, it dries and a tattoo is created.’
(Max Magop - DGB1-interview10-fs_mm - 5:50-6:03)

Data collection and corpus construction are on-going. Currently, the corpus comprises some 60,000 words with 66 speakers from all clans, and is undergoing further development as part of an on-going grant project (ELDP MDP0382). The text corpus has been analyzed from the perspective of lexical (Barth 2016, Barth Accepted) and syntactic variation (Barth and Anderson 2015), but no phonetic analysis to date has been conducted. We present here the first large-scale vowel analysis, capitalizing on the amount of data supplied by forced alignment.

4 Forced Alignment for Under-documented Languages

A key part of any phonetic description of a language, beyond capturing the structural phonemic elements of a language, is acknowledging and describing variability in the phonetic system. To be able to account for conditioning factors that affect phonetic realization, however, a relatively large amount of data from a representative sample of speakers is required. Increasingly in sociophonetic research, this is achieved by forced alignment—the automatic creation of time-aligned boundaries at the segment level from orthographic transcriptions aligned at the level of the utterance. As well as audio recordings and time-aligned, corresponding transcripts, this process requires resources that

[3]Data freely available in archives: catalog.paradisec.org.au/collections/DGB1, catalog.paradisec.org.au/collections/SocCog & elar.soas.ac.uk/Collection/MPI1194127

are not typically available for under-documented languages including a "dictionary", or a grapheme-to-phoneme (G2P) mapping, and an acoustic model. The creation of a force-aligned dataset for under-documented languages, then, appears to present a considerable challenge. However, as we demonstrate here, both G2P mapping and an acoustic model can be built from the same data that is to be aligned.

G2P mappings are one-to-one correspondences between the phonemes of a language and the graphemes that represent them (e.g., for Matukar Panau, *i* represents the phone [i], *y* represents [j], etc.). To force align Matukar Panau data, it was first necessary to create the G2P mappings. This was relatively straightforward for this language, thanks to its fairly phonemically transparent orthography (designed by community member Rudolf Raward at a workshop held by Living Tongues in 2010). However, there are some unresolved issues. As one example, transcribed *e* can be [e] or [ɛ] before nasals; to address this, all *e* tokens received the same G2P mapping, enabling the impact of phonological context to be examined in future work. In all, G2P mappings were created for 5,316 distinct word types comprising 33,632 total words spoken over 14.3 recording hours from 34 speakers.[4]

Second, it was necessary to build an acoustic model, which served as a blueprint for segment classification. One option is to apply a pre-trained model to new data. As is the case with most minority languages, no such pre-trained model exists for Matukar Panau. Past work on other minority languages has addressed this by applying an acoustic model built on the basis of one language to the language for which an acoustic model is lacking. For example, an English acoustic model built from Supreme Court Justice data has been applied through the application of the program FAVE (Rosenfelder et al. 2014) to the forced alignment of Cook Islands Māori (Coto-Solano and Nicholas 2018) and Bequia Creole (Walker and Meyerhoff In Press); and an Italian acoustic model available in MAUS (Kisler, Schiel and Sloetjes 2012) has been applied to Kriol (Jones et al. 2017). While this option has the benefit of ease of implementation and a community of users, it relies on matching phonemic inventories and orthographic systems of often completely unrelated languages and we do not know much the mismatch between acoustic models and target language affects the alignment.

Another option is to build an acoustic model on the basis of the data input to the aligner. This is possible through the train/align procedure available in some forced aligners. Train/align creates an acoustic model based entirely on the transcribed data for the language; thus, it is not biased by the phonemic inventory of another language but is optimized to suit the particular data being aligned. Comparisons of pre-trained and train/align models have found that an acoustic model trained on the input data yields better alignments for several minority languages (cf. Johnson et al. 2017, Johnson et al. 2018, using Prosodylab-Aligner (PL-A)).

Here we employ the train/align procedure available in the Montreal Forced Aligner (MFA, McAullife et al. 2017, which is an update of PL-A). Unlike many other aligners in wide use in linguistics which use the HTK toolkit (Young et al. 2009), MFA uses the more recent and powerful Kaldi (Povey et al. 2011). A major difference is that in using Kaldi, MFA employs both monophone and triphone models, rather than just monophone models as is typical of other aligners. During alignment, triphone models account for the phonological segments both preceding and following the vowel, as well as the feature profile of that segment (McAuliffe et al. 2017:498). This produces alignments that are more accurate and robust than those produced by aligners which employ HTK (cf. Gonzalez, Grama and Travis 2018), making MFA an excellent option for languages without existing acoustic models.

5 Optimizing Force-aligned Data

In order to maximize the data available, we applied an algorithm that improved the output from the forced-alignment post hoc by using recursion (see Gonzalez et al. 2018 for details). This recursive algorithm works by breaking the audio data and corresponding alignments into one-minute chunks, which are analyzed multiple times. With every new one-minute chunk that is analyzed, information

[4] The remaining data from the larger text corpus was either not available at the time of analysis or created errors and was discarded.

is added, re-evaluated, and used to refine boundaries that had been previously placed in existing data chunks. Multiple evaluations of the same data facilitate the best possible boundary placement, given the available data (cf. Moreno et al. 1998).

This post-hoc recursive algorithm proved effective in improving the forced alignment, seen in the higher Overlap Rate, or the proportion of force-aligned boundaries that approximate boundaries produced by a trained phonetician. As depicted in Figure 1, as more data chunks are processed by the recursive algorithm, the Overlap Rate between intervals established by a human coder and those established by the forced aligner increases, particularly sharply up to the 35-minute mark. Here, with an Overlap Rate of 0.67, the alignment reaches a level comparable to that attained for English data, as reported in Fromont and Watson (2016:426). In contrast, the non-recursive (i.e., linear) procedure yields more fluctuations as a function of the amount of data processed. The recursive method thus produces optimal forced alignment results (despite imperfect input data) that allow us to proceed with sociophonetic analysis.

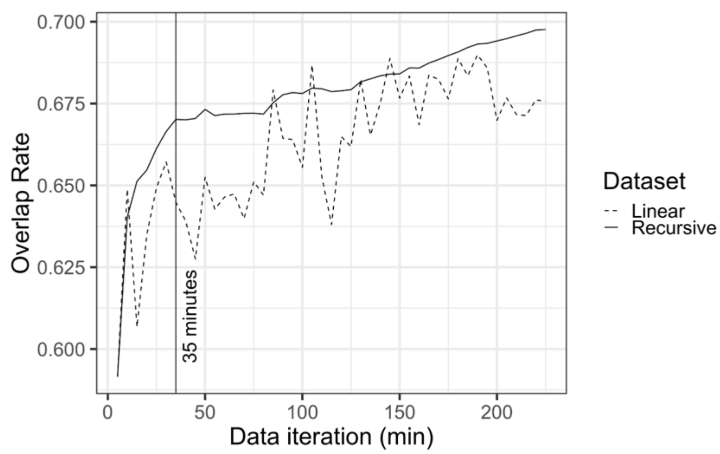

Figure 1. Overlap Rate Differences between Recursive and Linear approaches (from Gonzalez et al. 2018:147).

6 A First Sociophonetic Analysis of Matukar Panau

Forced alignment produced a total of 68,785 vowel tokens from which we were able to extract formant information. We followed standard procedure to account for the effect outliers might have on the data, and implemented a post-hoc filter at three standard deviations, calculated for each speaker's vowel category. This filtering removed approximately 3% of the total number of tokens (in line with the reduction in data size using a similar standard deviation filter reported in Foulkes et al. 2018:4, reporting on Hughes 2014, for English). The resultant Matukar Panau phonetic corpus comprises 66,942 vowel tokens, of which the most common vowel is /a/, comprising nearly half of the vowel tokens. Many of these tokens, however, are particularly short, and are unlikely to yield accurate formant measurements. As a result, vowels of less than 50ms were filtered out in an effort to control for the impact extremely reduced vowels might have on realizations. This removed 18,700 tokens (nearly 30% of the total number of vowels). Token numbers for vowels at the separate stages of data processing can be found in Table 1.

We cannot present a description of Matukar Panau phonological structure, as little is known about this. But as an overview of the data distribution, we note that in the overall data (prior to any filtering, and not controlling for word boundaries), a segment preceding a vowel tends to be an obstruent (47% of vowel tokens), while a segment following a vowel is most likely to be either a nasal (36%) or an obstruent (29%) (cf. Example (1)).

Vowel	Stage 1: Raw force-aligned data	Stage 2: Filtered at 3 std. dev.	Stage 3: Isolated tokens > 50ms
/i/	15,505	15,105	10,039
/e/	6,849	6,615	5,589
/u/	6,044	5,859	4,023
/o/	8,875	8,678	7,105
/a/	31,512	30,685	21,486
Total	68,785	66,942	48,242

Table 1. Vowels from the Matukar Panau corpus at each stage of data processing.

We now turn to consider social effects. To account for vocal tract length, vowels were normalized using the Lobanov (1971) method. Normalization was applied only to vowels which were greater than 50ms in duration. Analysis was conducted using R (R Core Team 2018) and plots generated using ggplot2 version 3.1.0 (Wickham et al. 2018). In order to more effectively characterize variation along social lines, we further restrict our sample to three cardinal vowels /i, u, a/ which do not occur adjacent to nasal consonants. This was done both to mitigate the well-known effect of nasal consonants on adjacent vowels, and because the phonemic status of /e, o/ as opposed to /ɛ, ɔ/ is as yet unclear in Matukar Panau. Even with these restrictions, thanks to the automated methods of alignment, we are left with 2,744 tokens of /i/, 1,200 tokens of /u/, and 6,742 tokens of /a/.

With a corpus of this size, we can begin to investigate the impact social factors play on vowel realizations. Clan, although a major organizing axis for cultures in various parts of the world, is a relatively underexplored sociolinguistic variable. One study that has taken account of clan (Stanford 2009) nevertheless found that it did play a role in variation in lexical tone realization among the Sui in China. Thus, here we investigate the difference between the major clan groups, the big vs. little brother sub-groups within each clan group (Bantibun: Bantibun & Binganen, Tamaten Honen: Tamaten Honen & Tamaten Painen), and for the offshoot clan (Kutur).

Before delving into clan, it is worthwhile considering the impact of the more widely studied social categories of age and gender. Table 2 shows the breakdown of the corpus for age, gender, clan head and clan type. Immediately, it is clear that detailed comparisons across clan head, type, gender and age together are impossible given the distribution of the data. It should also be clear that the sample is much larger for the Kutur offshoot group, and this is because the researcher (Barth) is based in the Kutur area and has spent more time collecting data from those that live in this area. Despite somewhat unbalanced numbers, some broad comparisons can be made across age and gender.

<table>
<tr><th rowspan="2">Clan Group</th><th rowspan="2">Clan Type</th><th rowspan="2">Gender</th><th colspan="2">Age</th><th rowspan="2">Total</th></tr>
<tr><th>Older</th><th>Younger</th></tr>
<tr><td rowspan="4">Bantibun</td><td rowspan="2">Big
(Bantibun)</td><td>F</td><td>2</td><td>5</td><td>7</td></tr>
<tr><td>M</td><td>1</td><td></td><td>1</td></tr>
<tr><td rowspan="2">Little
(Binganen)</td><td>F</td><td>2</td><td>3</td><td>5</td></tr>
<tr><td>M</td><td>1</td><td></td><td>1</td></tr>
<tr><td rowspan="7">Tamaten Honen</td><td rowspan="2">Big
(Tamaten Honen)</td><td>F</td><td></td><td>2</td><td>2</td></tr>
<tr><td>M</td><td></td><td>1</td><td>1</td></tr>
<tr><td rowspan="2">Little
(Tamaten Painen)</td><td>F</td><td></td><td>2</td><td>2</td></tr>
<tr><td>M</td><td>2</td><td></td><td>2</td></tr>
<tr><td rowspan="3">Offshoot
(Kutur)</td><td>F</td><td>1</td><td>2</td><td>3</td></tr>
<tr><td>M</td><td>2</td><td>1</td><td>3</td></tr>
<tr><td>Total</td><td>11</td><td>16</td><td>27</td></tr>
</table>

Table 2. Demographic breakdown across available social categories.

We first compare across age and gender, irrespective of clan, to test for possible overall changes in the community, as illustrated in Figure 2. Among the women, very little change appears to have taken place between older and younger speakers; both groups show largely identical means and distributions of these vowels, with the exception of a backer distribution of /i/ in older women. This suggests stability over time for the women. Men show somewhat more obvious differences, particularly in the positions of /i/ and /u/, which appear to have raised among the younger relative to the older speakers. While this is suggestive of a change in progress, it must be interpreted with some caution, as there are only two young men represented in the sample, and both are from the Tamaten Honen clan.

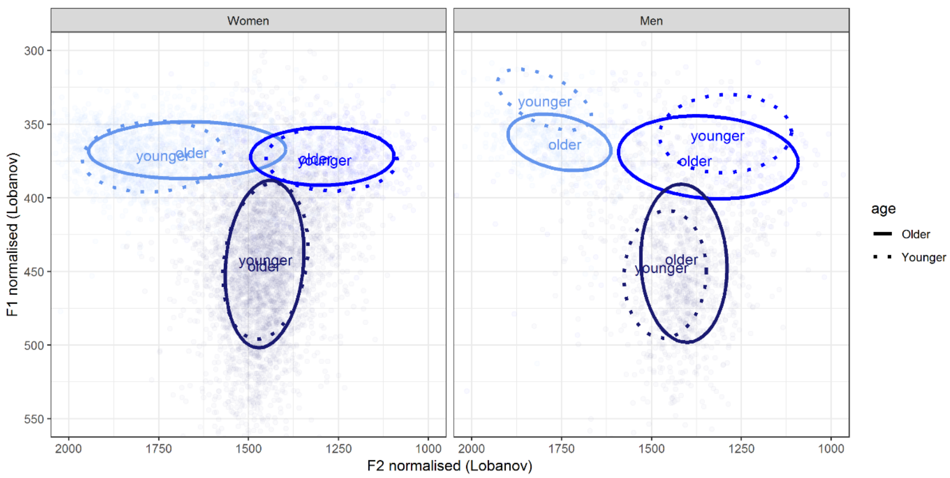

Figure 2. /i, u, a/ in normalized F1/F2 space for older (solid line) and younger (dotted line) women (left) and men (right); text represents group means (*n* = 10,686).

Moving away from age and gender, it appears that vowel realizations pattern reliably as a function of clan affiliation. Figure 3 shows the distributions of /i, u, a/ across major clan groups (Bantibun vs. Tamaten Honen) and subgroups (big vs. little vs. offshoot). There is a visible difference in the distribution of vowel realization by clan. The Bantibun-Binganen clan group shows relatively few within-clan differences, with nearly identical means and distributions for /u/ and /a/; /i/ appears to show a more retracted mean and distribution in Binganen speakers than Bantibun speakers, but this difference may be due to intersections with age or gender. By contrast, we observe more differences in Tamaten Honen. The little brother and offshoot sub-groups pattern more similarly to each other than either does to the big brother group; both the little brother and offshoot sub-groups have a marginally fronter /i/, lower /u/ and higher /a/ than the big brother group.

The starker differences for the Tamaten clan group as compared to the Bantibun clan group is perhaps reflective of the historical relationship between the offshoot group, and the little brother group from which it split. The greater differentiation could stem from groups wanting to avoid similarity with another group as well as greater physical distance and reduced contact after the schism, and/or coincidental individual speaker variation. Although the history of the clan schism is murky, it is clear that around 50 years ago, a family group left their Tamaten clan area in the main part of the village and moved some 500 meters away to what was previously part of a Catholic missionary coconut plantation. Moving created some physical distance, but this may have also been accompanied by increased social distance and a reduced level of contact. This potentially allowed for the development of a small change in vowel production, or there may have been more active distancing and change by either Kutur or other Tamaten members.

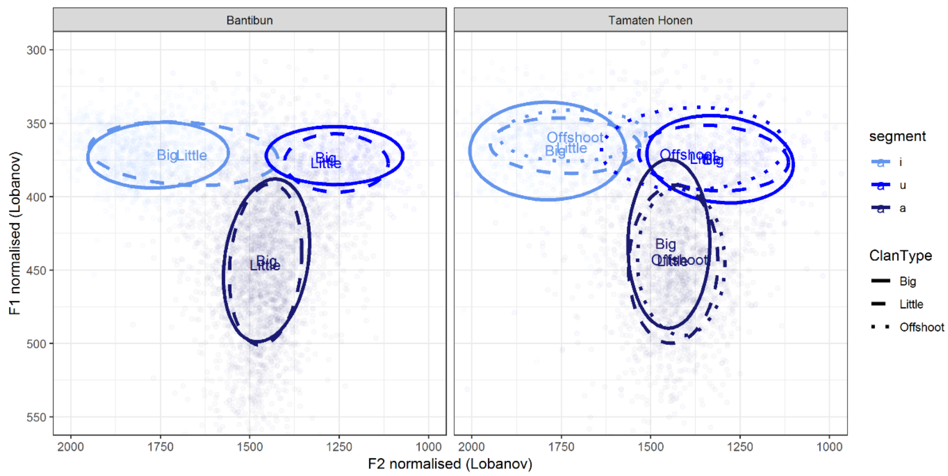

Figure 3. /i, u, a/ in normalized F1/F2 space for big brother (solid line), little brother (dashed line), and offshoot (dotted line) clan subgroups for Bantibun (left) and Tamaten Honen (right) clan heads; text represents group means.

It is also worth noting that researchers are not independent of community social networks. As stated earlier, the primary documentary linguist is hosted by a family in Kutur, hence the better representation of Kutur participants in the corpus. Some community members volunteer to participate in documentary recordings, and other community members are encouraged to do documentary recordings by her host family. She has had less contact with and less chance to document other Tamaten speakers than Bantibun speakers, itself perhaps due to lingering effects of the historical division. Future documentary work and greater inclusion of Tamaten Honen and Tamaten Painen clan members will allow further testing of how stable this variation is.

7 Conclusion

This work provides strong indications that the realization of cardinal vowels in Matukar Panau is socially conditioned, and that clan, as well as a major axis for community organization, is also an important factor in linguistic variation. The role of clan will best be understood in future work by taking a full account of its impact in relation to a range of predictors that may be affecting vowel realization, such as gender and age considered here, but also social networks, stylistic variation and identity, levels of bilingualism, as well as additional phonological conditioning. It is through the tools of forced-alignment that such work will be possible, providing sufficient tokens for analysis to meaningfully probe this full range of predictors.

References

Anderson, Gregory. D. S., Danielle Barth, and Kadagoi Rawad Forepiso. 2015. The Matukar Panau online talking dictionary: collective elicitation and collaborative documentation. In *Language documentation and cultural practices in the Austronesian world: papers from 12-ICAL*, Volume 4, eds. I W. Arka, N. L. N. S. Malini and I. A. M. Puspani, 111-126.

Barth, Danielle. 2016. Construction of expert identity in Matukar Panau. Paper presented at NWAV– AP4, National Chung Cheng University.

Barth, Danielle. To Appear. Variation in Matukar Panau kinship terminology. *Asia Pacific Language Variation*.

Barth, Danielle, and Gregory. D. S. Anderson. 2015. Directional constructions in Matukar Panau. *Oceanic Linguistics* 54:206-239.

Brognaux, Sandrine, Sophie Roekhaut, S., Thomas Drugman, and Richard Beaufort. 2012. Automatic phone alignment: A comparison between speaker- independent models and models trained on the corpus to align. In *Proceedings of the 8th International Conference on NLP*.

Carnegie Mellon University. 1993-2014. The CMU pronunciation dictionary. http://www.speech.cs.cmu.edu/cgi-bin/cmudict.

Coto-Solano, Rolando and Sally Nicholas. 2018. Using untrained forced alignment to study variation of glottalization in Cook Islands Māori. Paper presented at NWAV-AP5, University of Queensland.

DiCanio, Christian T., Hosung Nam, Douglas H. Whalen., H. Timothy Bunnell, Jonathan Amith, and Rey Castillo. 2012. Assessing agreement level between forced alignment models with data from endangered language documentation corpora. In *INTERSPEECH-2012*.

Foulkes, Paul, Gerry Docherty, Stefanie Shattuck Hufnagel, and Vincent Hughes. 2018. Three steps forward for predictability. Consideration of methodological robustness, indexical and prosodic factors, and replication in the laboratory. *Linguistics Vanguard* 42:1-11.

Fromont, Robert, and Kevin Watson. 2016. Factors influencing automatic segmental alignment of sociophonetic corpora. *Corpora* 11:401-431.

Goldman, Jean-Philippe, and Sandra Schwab. 2011. EasyAlign Spanish: an (semi-)automatic tool under Praat. In *Procedings of V Congreso de Fonética Experimental 2011*.

Gonzalez, Simón, James Grama, and Catherine Travis. 2018. Comparing the accuracy of forced-aligners for sociolinguistic research. Poster presented at CoEDL Fest, University of Melbourne.

Gonzalez, Simón, Catherine Travis, James Grama, Danielle Barth and Sunkulp Ananthanarayan. 2018. Recursive forced alignment: A test on a minority language. *Proceedings of 17th Speech Science and Technology Conference*.

Harrison, K. David, Gregory D. S. Anderson, and Danielle Barth**.** 2010-2012. Matukar-English online talking dictionary. http://matukar.swarthmore.edu/.

Hughes, Vincent. 2014. The Definition of the Relevant Population and the Collection of Data for Likelihood Ratio-based Forensic Voice Comparison. Doctoral dissertation, University of York.

Johnson, Lisa M., Marianna Di Paolo, Adrian Bell and Carter Holt. 2017. Forced alignment for understudied language varieties. Paper presented at ICLDC5, University of Hawai'i.

Johnson, Lisa M., Marianna Di Paolo, and Adrian Bell. 2018. Forced alignment for understudied language varieties: Testing Prosodylab-Aligner with Tongan Data. *Language Documentation and Conservation* 12:80-123.

Jones, Caroline, Katherine Demuth, Weicong Li, and Andre Almeida. 2017. Vowels in the Barunga variety of North Australian Kriol. *INTERSPEECH-2017*.

Kaspruś, Aloys. 1942. The languages of the Mugil District, NE-New Guinea. *Anthropos* 37:711-778.

Kisler, Thomas, Florian Schielk and Han Sloetjes. 2012. Signal processing via web services: The use case WebMAUS. Paper presented at the Digital Humanities Conference 2012, University of Hamburg.

Labov, William, Ingrid Rosenfelder, and Josef Fruehwald. 2013. One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language* 89:30-65.

Lausberg, Hedda and Han Sloetjes. 2009. Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, Instruments, & Computers (Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. http://tla.mpi.nl/tools/tla-tools/elan/)*. 41(3), 841-849.

Lobanov, Boris. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America* 49:606-608.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger, 2017. Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. *INTERSPEECH-2017*, 498-502

Milne, Peter. 2014. The Variable Pronunciations of Word-final Consonant Clusters in a Force Aligned Corpus of Spoken French. Doctoral dissertation, University of Ottawa.

Moreno, Pedro J., Chris Joerg, Jean-Manuel Van Thong, and Oren Glickman. 1998. A recursive algorithm for the forced alignment of very long audio segments. In *International Conference on Spoken Language Processing-1998*.

Povey, Daniel, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, Jan Silovsky. Georg Stemmer and Karel Veselý. 2011. *The Kaldi speech recognition toolkit*. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society.

R Core Team. 2018. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, http://www.R-project.org.

Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini, Scott Seyfarth, Kyle Gorman, Hiliary Prichard and Jiahong Yuan. 2014. FAVE (Forced Alignment and Vowel Extraction) Program Suite v1.2.2 10.5281/zenodo.22281.

Schmid, Monika, and Barbara Köpke, B. 2007. Bilingualism and attrition. In *Language attrition: Theoretical perspectives*, eds. B. Köpke, M. S. Schmid, M. Keijzer, and S. Dostert, 1-7. Amsterdam: John Benjamins.

Stanford, James N. 2009. Clan as a sociolinguistic variable: Three approaches to Sui clans. In *Variation in indigenous minority languages*, eds. J. Stanford and D. Preston, D., 463-84. Amsterdam: John Benjamins.

Walker, James, and Miram Meyerhoff. In Press. Pivots of the Caribbean? Low-back vowels in Eastern Caribbean English. *Linguistics*.

Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pederson, Kohske Takahashi, Claus Wilke, Kara Woo, and Hiroaki Yutani. 2019. ggplot2: Creating elegant data visualisations using the grammar of graphics. R package version 3.2.0.

Young, Steve, Gunnar Evermann, Mark Gales, Thomas Hain, Dan Kershaw, Xunying Liu, Gareth Moore, Julian Odell, Dave Ollason, Daniel Povey, Valtcho Valtchev, and Phil Woodland. 2009. *The HTK book (for version 3.4)*. Cambridge: Cambridge University Engineering Department.

Yuan, Jiahong, and Mark Liberman. 2008. Speaker identification on the SCOTUS corpus. *Journal of the Acoustical Society of America* 123:9688-9690.

Z'Graggen, John A. 1969. Classificatory and typological studies in languages of the Madang district. *Pacific Linguistics* C-19. Canberra: Australian National University.

Centre of Excellence for the Dynamics of Language
Australian National University
Acton, ACT 2600
Australia
danielle.barth@anu.edu.au
james.grama@anu.edu.au
simon.gonzalez@anu.edu.au
catherine.travis@anu.edu.au