

Exploring automated formant analysis for comparative variationist study of Heritage Cantonese and English

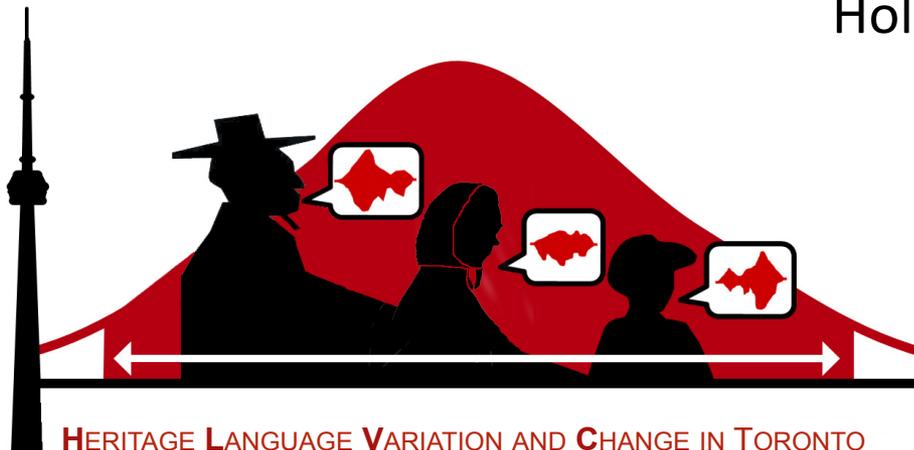


Naomi Cui¹, Minyi Zhu¹, Vina Law¹

Holman Tse² & Naomi Nagy¹

¹University of Toronto

²University of Pittsburgh



HERITAGE LANGUAGE VARIATION AND CHANGE IN TORONTO
[HTTP://PROJECTS.CHASS.UTORONTO.CA/NGN/HLVC](http://projects.chass.utoronto.ca/ngn/hlvc)

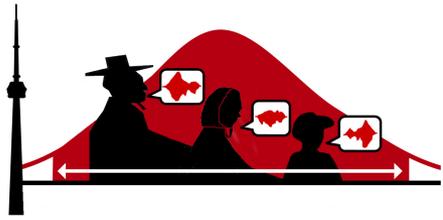


Social Sciences and Humanities
Research Council of Canada

Conseil de recherches en
sciences humaines du Canada



UNIVERSITY OF
TORONTO



What is the HLVC Project?

- Large-scale project investigating language use and change in heritage (non-official) languages spoken in Toronto.
- Goals
 - To **document and describe heritage languages** spoken by immigrants and 2 generations of their descendants
 - To **create a corpus** available for research on language change
 - To **push variationist research beyond its monolingually-oriented core** by focusing on heritage language use among multilingual speakers
 - To develop a framework for research on heritage languages and contact

A Sample of Previous HLVC Work

	Cantonese	Faetar	Italian	Korean	Russian	Ukrainian
VOT	✓		✓	✓	✓	✓
∅-subject	✓	✓	✓		✓	
Borrowing		✓				
Vowels	*					

* This presentation

Vowels

- Very well researched in sociolinguistics, but very little work on vowel variation and change in languages other than English.
- Large body of research has made possible the development of new technologies/techniques to make vowel analysis easier
 - Example: FAVE (Rosenfelder et al 2011)

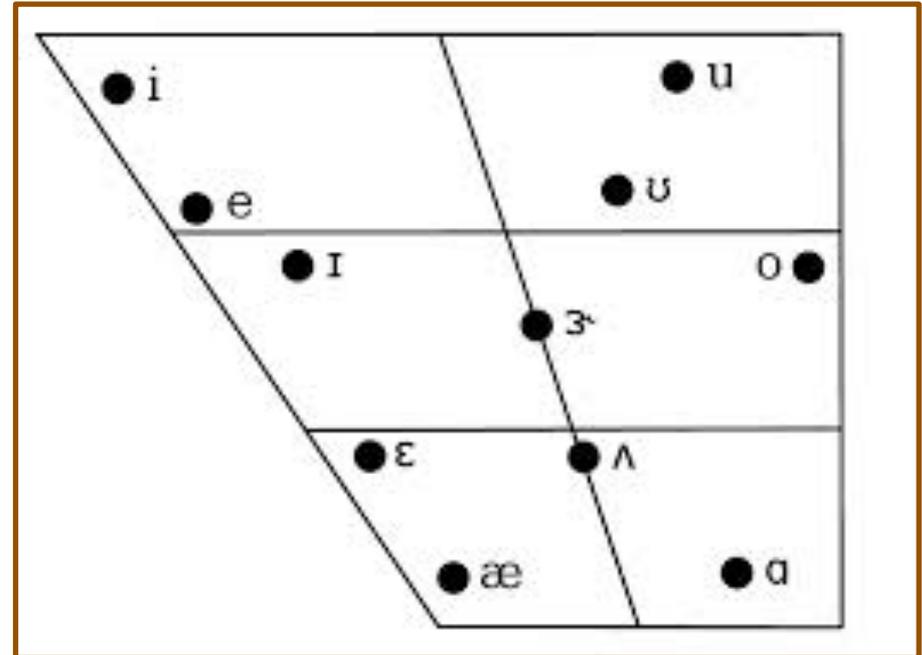


Image from Wikipedia

 **Forced Alignment & Vowel Extraction (FAVE)**
An online suite for automatic vowel analysis

Goals of Current Project

- To determine the extent to which the vowel systems of Cantonese and English may be mutually influencing each other in Toronto
- To extend the use of automated forced alignment and formant extraction as tools for the sociolinguistic study of contact-induced change in Heritage Cantonese.
 - Prosodylab-Aligner (Gorman et al 2011) to be adopted



Methodological Problems

- Large amount of data in HLVC Corpus (~40 speakers/language)
 - Manual formant measurements take a lot of time.
- FAVE designed to work only on English
- Could Prosodylab-Aligner be a viable alternative?



The language quilt

By Catherine Farley and Damian Lister/TORONTO STAR

English is still, by far, the first language across Greater Toronto. But strip away that blanket of dominance and a colourful patchwork emerges, showing where newcomers from around the world chose to settle. The map shows the most prevalent mother tongue after English in more than 1,000 neighbourhoods across the GTA, as revealed by a Star analysis of 2006 census data

Some discoveries

- English is the second language in 47 of the GTA's 1,076 census tracts
- English is in third place in 7 tracts in Agincourt, on Toronto's northern edge
- In 57 tracts, 70% or more of the population has a non-English mother tongue
- The preponderance of English as a mother tongue is 90% or more in 42 census tracts
- In 200 tracts, more than 30 distinct mother tongues are spoken by 15 or more people — the minimum number of speakers required for a language to count in the census
- In 13 tracts, there are more than 40 mother tongues

56% of the 5.4 million GTA residents count English as their mother tongue. Here are the next most prevalent:

Top 10 mother tongues (after English)

- Italian 3.5%
- Chinese (no language specified) 3.2%
- Cantonese 3.1%
- Punjabi 2.5%
- Portuguese 2%
- Spanish 2%
- Tagalog 1.9%
- Urdu 1.8%
- Tamil 1.7%
- Polish 1.6%

Behind the mosaic

Tight knit: The Greek community, comprising 43,200 speakers, is one of the GTA's most compact with Greek placing second behind English in 20 census tracts in central Toronto. Koreans, with 4,200 speakers, is one of the smallest communities, but it ranks No. 2 in one rural community in North Durham, ahead of the German and French contingents.

Loose knit: Mother tongues with wide distribution rarely get into second place. Although a sizable community in numbers (23,833), Hindi has no dominant census tracts and Hungarian (21,325) has just one. In contrast, Ukrainians, with 29,040 speakers, command 17 No. 2 positions.

In the shadows: Some languages are under-represented on the map because they are overshadowed by other dominant tongues. In parts of North Toronto and Markham, where Chinese and Cantonese rank 1 and 2, Mandarin is consistently third or fourth. Despite 63,140 speakers, Mandarin only achieves second place in 6 census tracts.

Mother tongue: First language learned in childhood and still understood. The 2006 Census identifies 116 distinct mother tongues spoken by at least 16 people in the GTA. **Census tracts:** A neighbourhood or rural area defined by geographic boundaries such as rivers, highways and streets, with typically 8,000 to 15,000 residents. ***Chinese (no language specified):** May include Cantonese, Mandarin, Hakka, Taiwanese, Fujian or Shanghaiese language

KEY TO MOTHER TONGUES
Number of speakers in the GTA

Western Europe, Americas

- English 3,629,955 (66%)
- Italian 189,778 (3.5%)
- Portuguese 160,258 (3%)
- Spanish 93,225 (2%)
- French 85,435 (1.6%)
- Greek 43,988 (0.8%)
- German 42,955 (0.8%)
- Dutch 15,460 (0.3%)

East Asia

- Chinese (no language specified) 170,439 (3.2%)
- Cantonese 167,395 (3.1%) (Hong Kong)
- Tagalog 104,485 (1.9%) (Philippines)
- Mandarin 63,140 (1.2%) (China)
- Korean 65,385 (1.2%)
- Vietnamese 45,736 (0.8%)

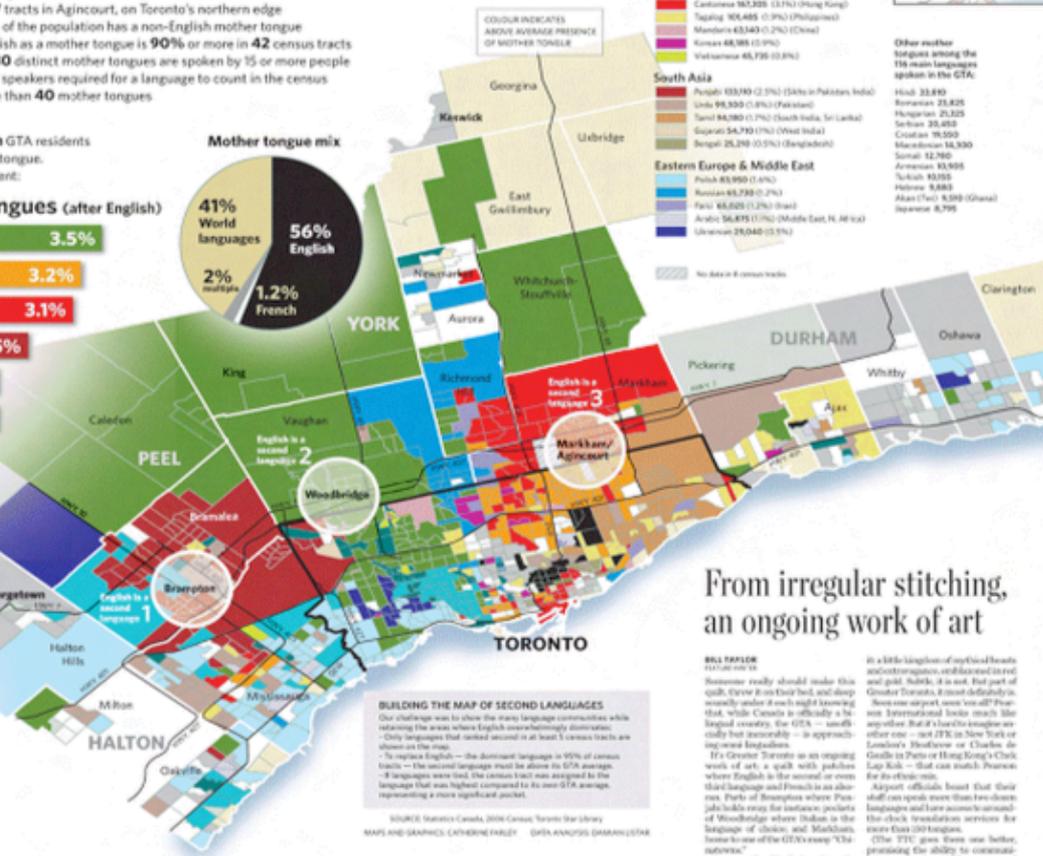
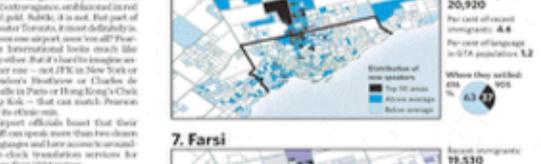
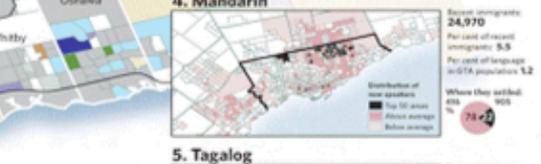
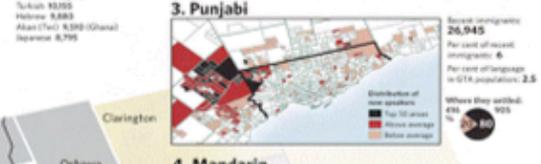
South Asia

- Punjabi 233,930 (4.2%) (India in Pakistan, India)
- Hindi 22,890
- Burmese 21,825
- Hungarian 21,325
- Sinhalese 20,405
- Croatian 19,500
- Macedonian 14,300
- Slovak 12,765
- Armenian 13,905
- Turkish 13,055
- Polish 12,685
- Alban (Thai) 9,590 (0.18%)
- Japanese 6,795

Eastern Europe & Middle East

- Polish 83,950 (1.6%)
- Russian 65,730 (1.2%)
- Farsi 65,000 (1.2%) (Iran)
- Ukrainian 29,040 (0.5%)

No data in 8 census tracts



BUILDING THE MAP OF SECOND LANGUAGES
Our challenge was to show the many language communities while retaining the areas where English overwhelmingly dominates. Only languages that ranked second in at least 5 census tracts are shown on the map. To reverse English — the dominant language in 90% of census tracts — the second language must be above its GTA average. If languages were tied, the census tract was assigned to the language that was highest compared to its own GTA average, representing a more significant pocket.

SOURCE: Statistics Canada, 2006 Census, Toronto Star Library
MAPS AND GRAPHICS: CATHERINE FARLEY DATA ANALYSIS: DAMIAN LISTER



From irregular stitching, an ongoing work of art

BILL FAYLOR
Illustrator

Business really should make this quilt. There it is on their bed, and sleep soundly under it each night knowing that, while Canada is officially a bi-lingual country, the GTA — a word that only has its own meaning — is approaching new languages.

If a Greater Toronto is an ongoing work of art, a quilt with patches where English is the second or even third language and French is an after-thought, it's a quilt that's been sewn in the language of choice and Markham, home to one of the GTA's most "suburban" areas.

None of the North American cities that can boast multiple Chateaus, it's not for nothing that the United Nations has called Toronto the world's most ethnically diverse city. For the most part the contemporary stitching is subtle. It's often less than apparent that English has been overthrown in the lingua franca.

Perhaps a handful of areas with an irregular edge, identifiable to you, represent sewing under the fringe. A church with its name in a script unknown to you. The owner's name on a street corner or park bench.

Observing it, a neighborhood probably feels different from your own — whatever language you speak that be chosen.

That's how a Design City in Markham, said to be the largest Chinese mall in North America, is celebrating

its 18th birthday of applied beauty and extravagance, elaborated in red and gold. Fabric, it is not. But part of Greater Toronto, it most definitely is. There one cannot see how all these international lines match like apples. But if it has the numbers — other one — not TTC in New York or London's Heathrow or Charles de Gaulle in Paris or Hong Kong's Chek Lap Kok — that can match. Perhaps it is to choose one.

Airport officials boast that their staff can speak more than five dozen languages and have access to around the clock translation services for more than 200 languages.

Other TTC goes. There are now better, meaning the ability to communicate in more than 170 languages and dialects.

From Mississauga to Delta, ethnologists say that there are some 6,800 languages spoken around the world. It's not hard to imagine 6.8% of them in Toronto and perhaps 90% in the GTA.

Based in the central area of Province on a busy Saturday and you will see people from, literally, everywhere. Coming home.

How being where the best is — a combination of neighbourhoods, languages spoken and where the patterns and shagwags flow and merge into a seamless tapestry of life.

You can say that any way you want to assist in the language of your choice. It means the same.

Or you could organize a quilting bee and make the world's best.

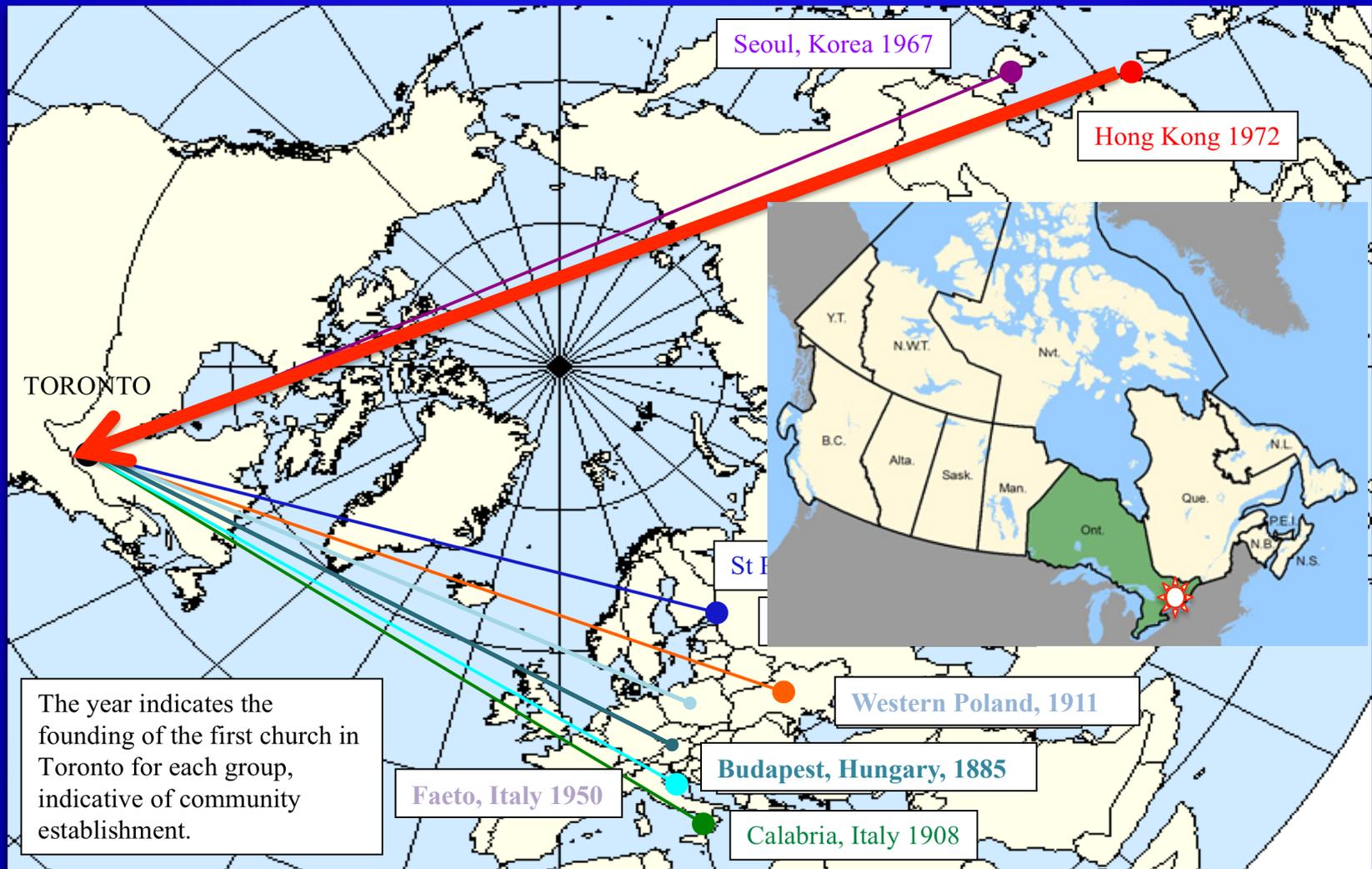
Contrasting demographics

Language	MT speakers (2011 Census)	Ethnic Origin (2006 Census)	Est. in TO	Speakers come from
Cantonese	170,000 ⁺	537,000	1951	Hong Kong
Italian	166,000	466,000	1908	Calabria
Russian	78,000	58,505	1916	St. Petersburg, Moscow
Korean	51,000	55,000	1967	Seoul
Ukrainian	26,000	122,000	1913	Lviv
Faetar	<100?	300?	1950	Faeto, Celle di St. Vito (Apulia Italy)

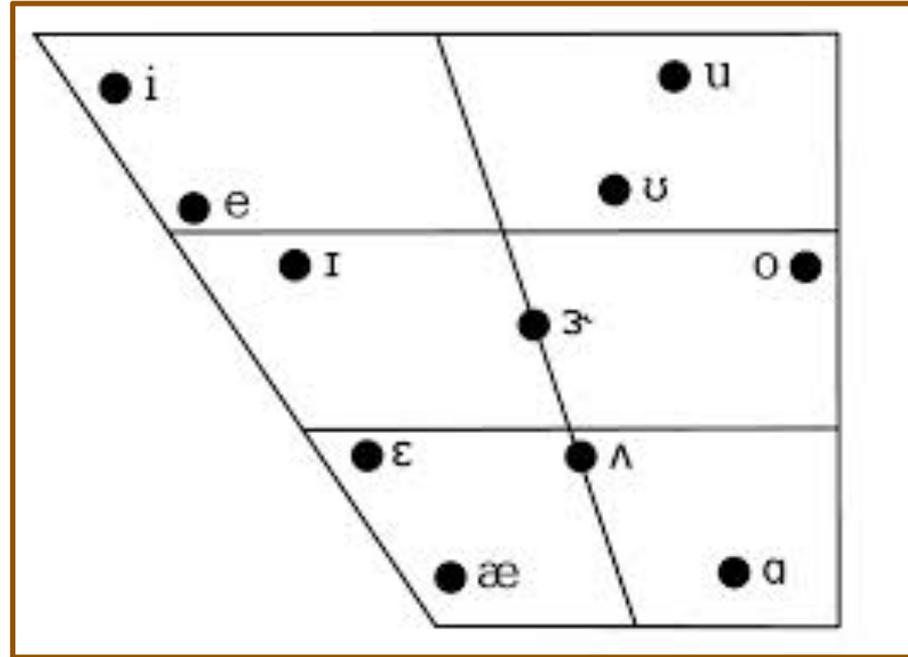
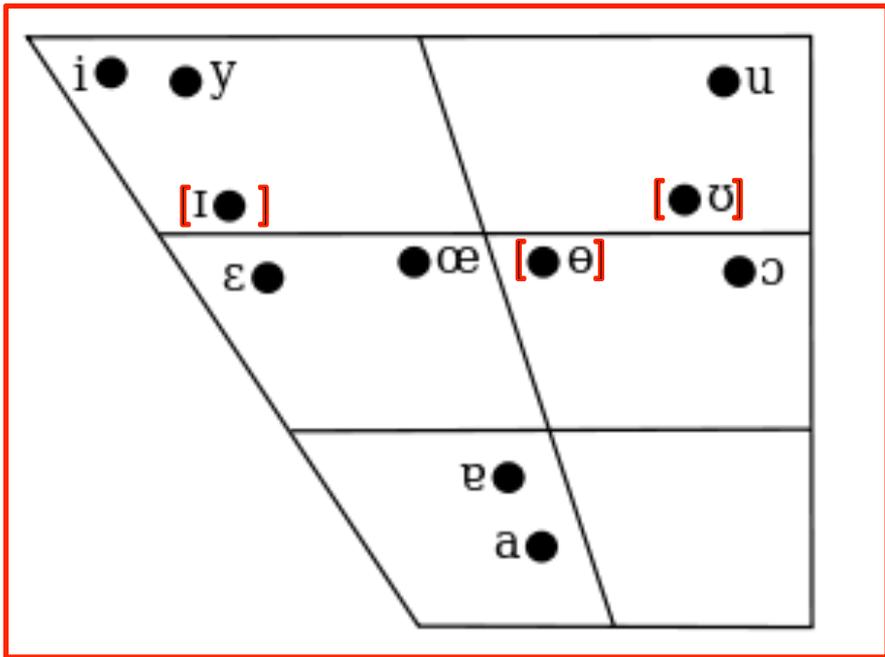
www40.statcan.ca/l01/cst01/demo12c-eng.htm;

www12.statcan.gc.ca/census-recensement/2011/dp-pd/prof/index.cfm?Lang=E

Heritage Language Variation and Change



Cantonese vs. English Vowel Space



Images from Wikipedia

Allophonic lowering of /i/ before velars
(Yue-Hashimoto 1972)

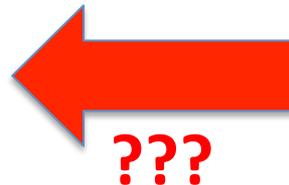
si1, /si1/, 詩, 'poem'

sik1, [sɪk1], 識, 'to know'

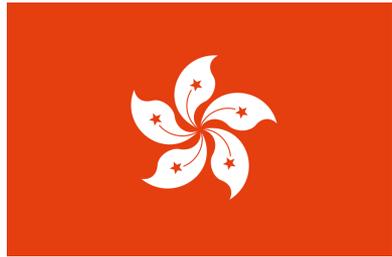
Similar Canadian English Vowels

see, /si/

sick, /sɪk/



Expected outcome



1st

2nd



Heritage Language / Culture

English/Canadian

Data

- Two sets of hour-long sociolinguistic interviews from 2 generations of speakers identified as Hong Kong Chinese and who claim Cantonese as a heritage language
 - Not from the same speakers, however.

Interviews in English from the Contact in the City Corpus (CinC) (Hoffman and Walker 2010)

Interviews in Cantonese from the HLVC Corpus (Nagy 2009, 2011)

“My parents came to Toronto in 1972.”



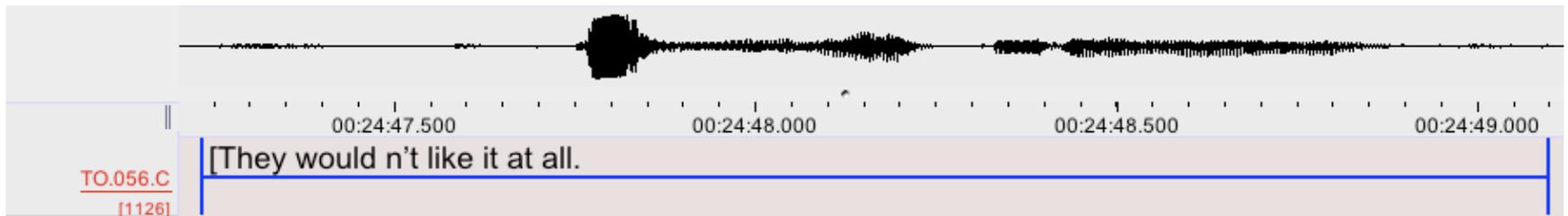
“Ngo5 fu6 mou5 yat1 gau2 cat1 yi6 lin4 lei4 dou3 do1 leon1 do1.”

Speaker Sample

Generation	Sex	CANTONESE	ENGLISH
1 (Ages: 42-82)	Male	C1M62A C1M59A	TO.035 TO.038
	Female	C1F78A C1F54A C1F82A	TO.030 TO.037 TO.039
	2 (Ages: 16-44)	Male	C2M44A
	Female	C2F16A C2F21B	TO.031 TO.056
Total		N=8	N=8

Methods - English Data

1. Sentence-level time alignment (manual) using ELAN



2. Word- and phoneme-level time alignment (automated) with FAVE
 - <http://fave.ling.upenn.edu>

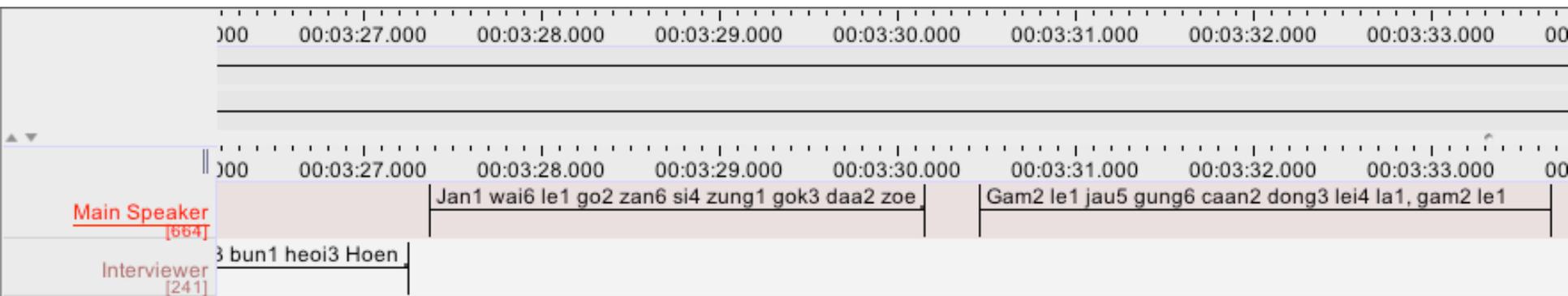
Prosodylab-Aligner (Gorman 2011 et al)



- A Python script used to perform text to audio speech alignment
- Supports training on arbitrary data
 - → With any input from language X, can be trained to deal with acoustic data from language X
- Requirements
 - At least a total of one hour of audio (.wav file in chunks OK)
 - Matching .lab files (.txt files readable by Prosodylab-Aligner) for each .wav file
 - A customized dictionary

Methods – Cantonese Data

1. Interviews transcribed by native speakers of Cantonese using Jyutping Romanization in ELAN
 - Manual sentence-level alignment



2. To create input readable by Prosodylab-Aligner, PRAAT script used to create smaller .wav files with matching .txt files for each annotation.

PRAAT Script

00:03:27.000 00:03:28.000 00:03:29.000 00:03:30.000 00:03:31.000 00:03:32.000 00:03:33.000

Main Speaker [664] Jan1 wai6 le1 go2 zan6 si4 zung1 gok3 daa2 zoe | Gam2 le1 jau5 gung6 caan2 dong3 lei4 la1, gam2 le1

Interviewer [241] 3 bun1 heoi3 Hoen |



C1F54A_IV_2074.wav



Jan1 wai6 le1 go2 zan6 si4 zung1 gok3 daa2 zoeng3.

Translation: "Because at that time, China was at war."



C1F54A_IV_2105.wav

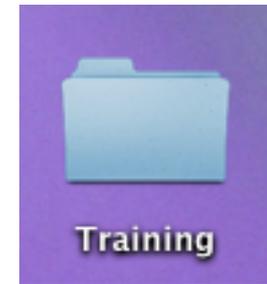


Gam2 le1 jau5 gung6 caan2 dong3 lei4 la1, gam2 le1

Translation: "And then the Communist Party came, and then ..."

Training and Evaluation

```
CAN_dict_4.txt
ZUNG3 Z U NG
ZUNG4 Z U NG
ZUNG5 Z U NG
ZUNG6 Z U NG
ZUT1 Z U T
ZUT2 Z U T
ZUT3 Z U T
ZUT4 Z U T
ZUT5 Z U T
ZUT6 Z U T
ZYU1 Z YU
ZYU2 Z YU
ZYU3 Z YU
ZYU4 Z YU
ZYU5 Z YU
ZYU6 Z YU
ZYUN1 Z YU N
ZYUN2 Z YU N
ZYUN3 Z YU N
ZYUN4 Z YU N
ZYUN5 Z YU N
ZYUN6 Z YU N
ZYUT1 Z YU T
ZYUT2 Z YU T
ZYUT3 Z YU T
ZYUT4 Z YU T
ZYUT5 Z YU T
ZYUT6 Z YU T
```



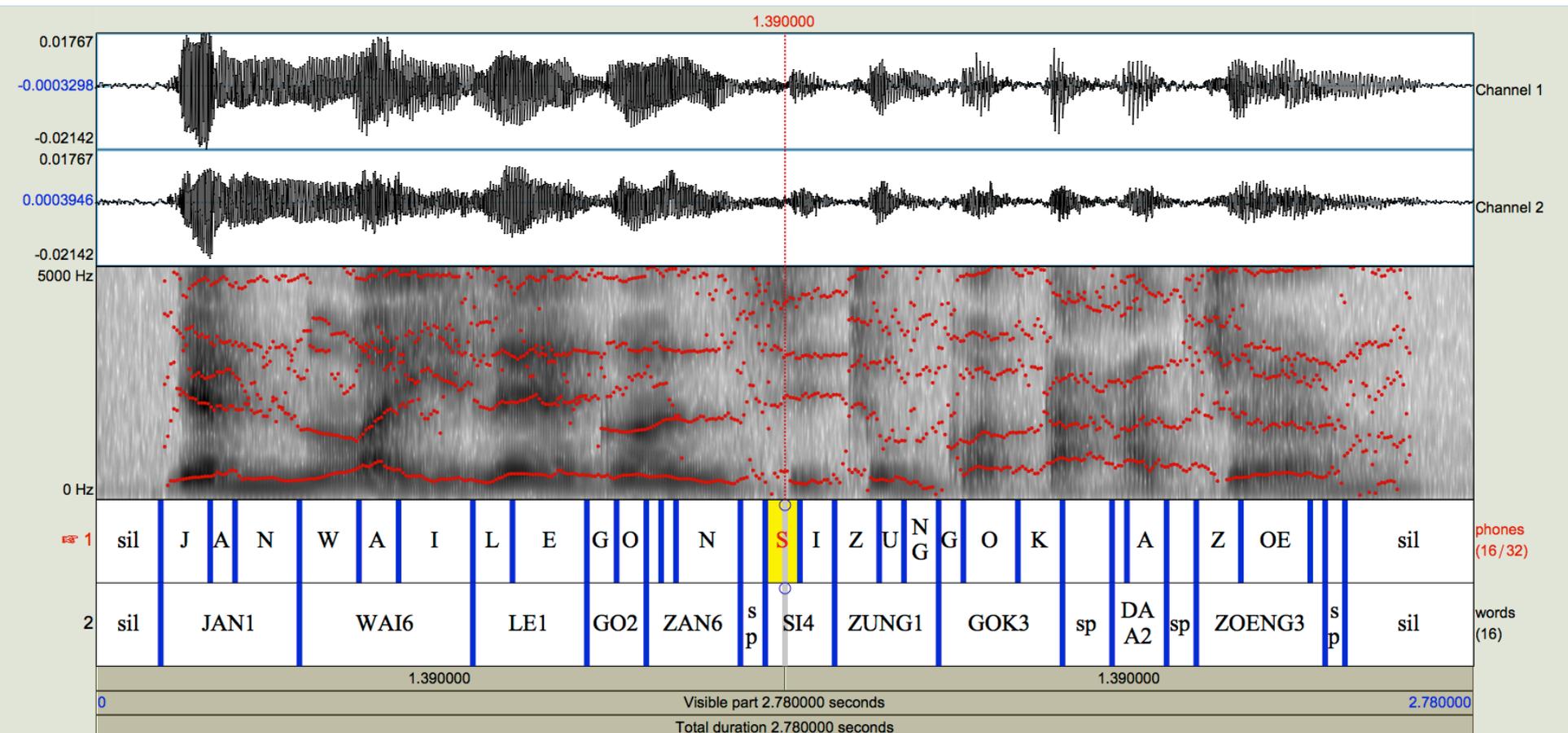
- .wav files and matching .lab files put in a Training directory
- Prosodylab-aligner uses Training directory and **dictionary** to build a training model



- **Prosodylab-aligner uses training model to evaluate the same files in the same directory**

Custom dictionary in the format of The CMU Pronouncing Dictionary

Textgrid Output of Prosodylab-Aligner



Another PRAAT script: formant extraction

- Formant information extracted from Prosodylab-Aligner generated Textgrids and matching .wav files using PRAAT script
- Output: Tab-delimited .txt file

	A	B	C	D	E	F	G	H	I	J
1	Speaker	File	Vowel	F1	F2	Word	Tone	Length		
2	C1F54A	10035	A	225	1851	HAI		6	0.14	
3	C1F54A	10035	I	314	2006	HAI		6	0.09	
4	C1F54A	10035	A	557	2024	AH		3	0.03	
5	C1F54A	10035	I	413	1988	TEOI		3	0.04	
6	C1F54A	10035	A	718	1678	JAU		1	0.06	
7	C1F54A	10035	U	523	1579	JAU		1	0.05	
8	C1F54A	10035	A	635	1921	GAI		3	0.08	
9	C1F54A	10035	I	471	2032	GAI		3	0.05	
10	C1F54A	10035	AA	739	1422	WAAK		6	0.04	
11	C1F54A	10161	I	422	1904	ZEOI		3	0.08	
12	C1F54A	10161	E	463	2301	LEI		2	0.07	
13	C1F54A	10161	I	372	2388	LEI		2	0.11	
14	C1F54A	10161	O	487	1148	GO		2	0.07	
15	C1F54A	10161	AA	650	1489	FAAN		3	0.08	
16	C1F54A	10161	U	412	2413	GUNG		1	0.16	
17	C1F54A	10161	A	832	1845	AH		3	0.16	
18	C1F54A	10161	I	345	1829	ZEOI		3	0.09	
19	C1F54A	10161	U	463	1112	ZUNG		1	0.13	
20	C1F54A	10161	I	433	2340	JI		3	0.05	
21	C1F54A	10161	O	413	1440	ZOU		6	0.03	
22	C1F54A	10161	U	423	1183	ZOU		6	0.12	
23	C1F54A	10161	I	392	1889	ZEOI		3	0.1	
24	C1F54A	10161	E	418	2104	LEI		2	0.03	
25	C1F54A	10161	I	387	2228	LEI		2	0.11	

Vowel Normalization

NORM

The Vowel Normalization and Plotting Suite



- <http://ncslaap.lib.ncsu.edu/tools/norm/norm1.php>
- Labov ANAE (Vowel Extrinsic) method used

1. Select the vowel data file: pre-NORM_template_CAN_8_speakers_GEN1and2_only_2014-05-26.txt
Ensure that the file is in a tab-delimited text format and that it matches the format of the NORM Template; see [How to Use NORM](#) for help. [\[Download Template\]](#)

2. Select result type:

3. Select normalization methods:

You can select multiple methods by holding the **ctrl** or **cmd** keys while clicking.

For information about each method, visit the [Methods](#) page.
Not all methods are equally appropriate for all vowel datasets. Note that, even though the primary component(s) for each method are listed, a number of the methods are slightly modified from their original, published versions. Specifications of these modifications are also described on the [Methods](#) page.

4. Select options:

Web Page Layout: Only relevant if more than one method selected.

Normalization & Processing:

Include F3 in Processing:
Only relevant for Nearey and Labov methods. Bark difference requires F3 values. Labov and Watt & Fabricius are not implemented to use F3. Note that including F3 for Nearey and Labov methods changes the results for values of F1 and F2; it is not just a matter of whether or not you want F3 values normalized.

Scale results:
Only relevant for Labov, Nearey and Watt & Fabricius methods, which otherwise result in non-hz or non-Barks values. Important: we do not recommend scaling. Make sure that you have read the [About Scaling](#) section.

Plotting:

Plot:

Plot Standard Dev:

Plot Labels:

Plot Legend:

Plot Dot Shapes:

Plot Dot Sizes:

Inline Plot Size:
This sets the size of the plots that appear inline on the results web page.

5. Click the button to upload your file for normalization:

Plot title:

Plotting Colors:
You can override NORM's plotting colors. To do so, enter a list of colors in the below form separated by commas. Colors can be entered using hex codes (e.g., #ff0000) or valid color names recognized by R (e.g., darkgrey). Visit a website like [this](#) or use the `colors()` function in R to view recognized colors. If you do not enter enough colors for your plot, the colors will be recycled. [Click here](#) for the default colors to be inserted in the form. [Click here](#) for an example of color names to be inserted in the form.

Prepping for R-Brul

- Tab-delimited .txt file generated by NORM with normalized values for vowel formants
- New columns added for variables
- Ready for statistical analysis with R-brul (Johnson)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Speaker	Vowel	Word	F1	F2	Generation	Sex	Age	Preceding	Following	Velar	Syllable	Tone
2	C1F54A	AA	BAAK6	857.44	1692.9	1	F	54	B	K	Yes	Closed	6
3	C1F54A	AA	BAAK6	815.52	1565.2	1	F	54	B	K	Yes	Closed	6
4	C1F54A	AA	BAAN1	821.51	1667	1	F	54	B	N	No	Closed	1
5	C1F54A	AA	BAAN1	818.51	1721.9	1	F	54	B	N	No	Closed	1
6	C1F54A	AA	BAAN3	815.52	965.24	1	F	54	B	N	No	Closed	3
7	C1F54A	AA	BAAN1	651.81	1606.1	1	F	54	B	N	No	Closed	1
8	C1F54A	AA	BAAN1	851.45	1546.2	1	F	54	B	N	No	Closed	1
9	C1F54A	AA	BAAN1	963.25	1518.2	1	F	54	B	N	No	Closed	1
10	C1F54A	AA	BAAN1	783.57	1394.5	1	F	54	B	N	No	Closed	1

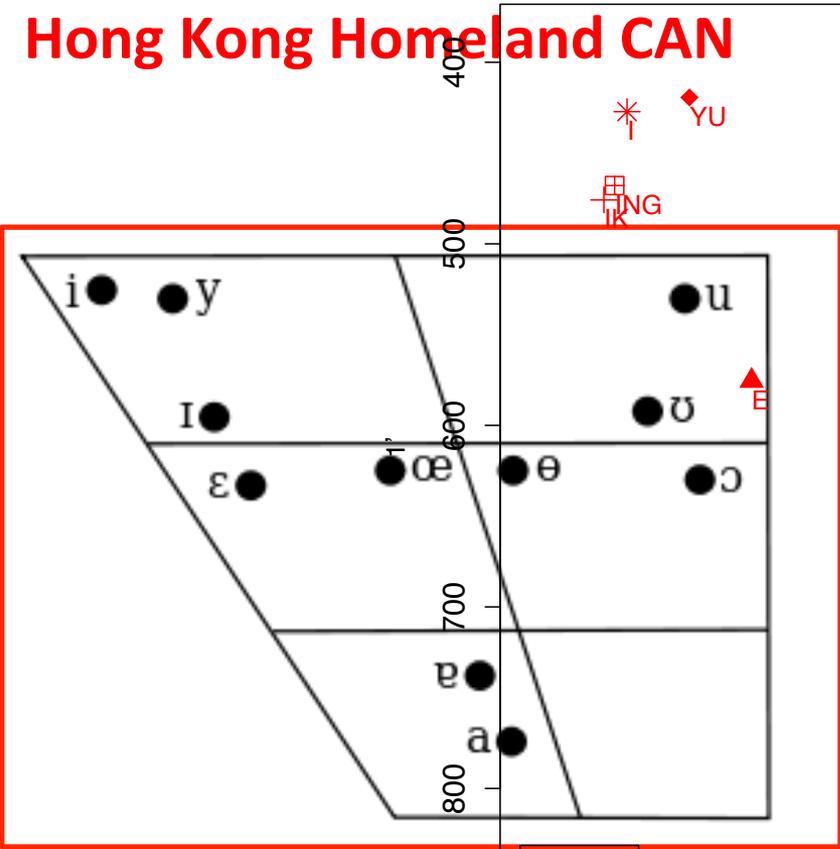
Variables of Interest

- External Factors
 - Generation
 - Gender
 - Age
- Internal Factors
 - Following Segment
 - Tone

Cantonese Vowel Charts

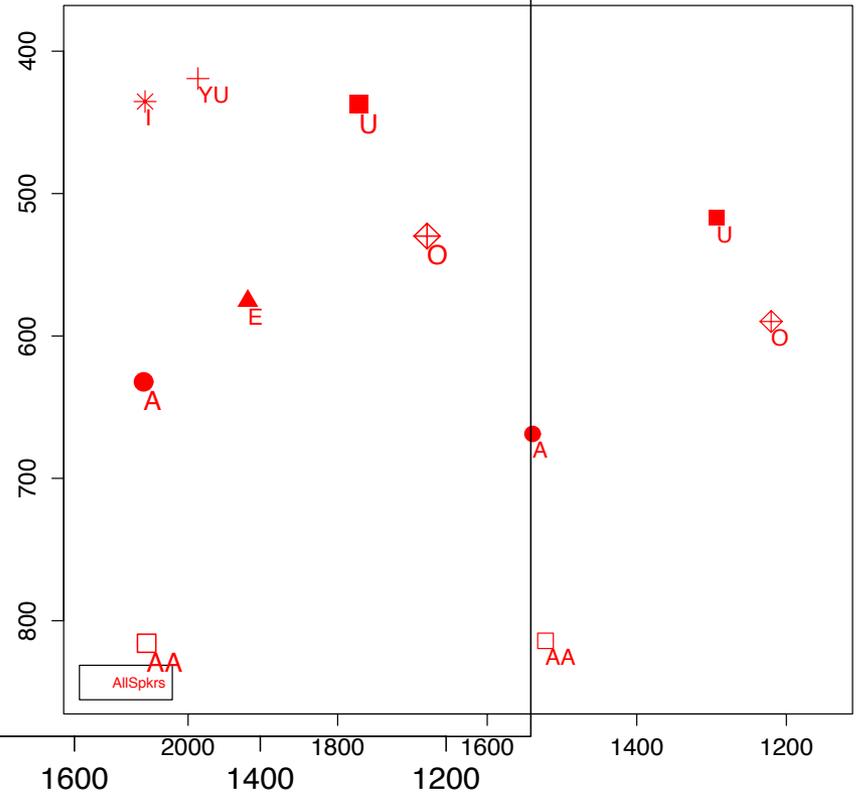
Toronto CAN (8 speakers), Labov ANAE (speaker extrinsic)

Hong Kong Homeland CAN



AllSpkrs image from Wikipedia

Toronto CAN (8 speakers), Labov ANAE (speaker extrinsic)

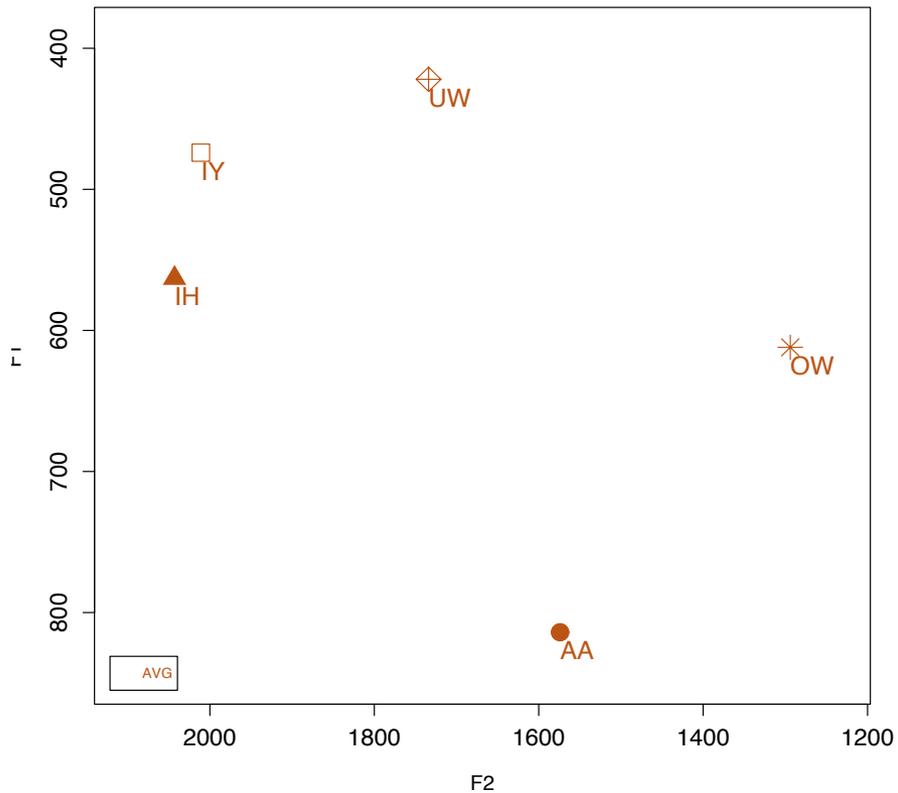


AllSpkrs

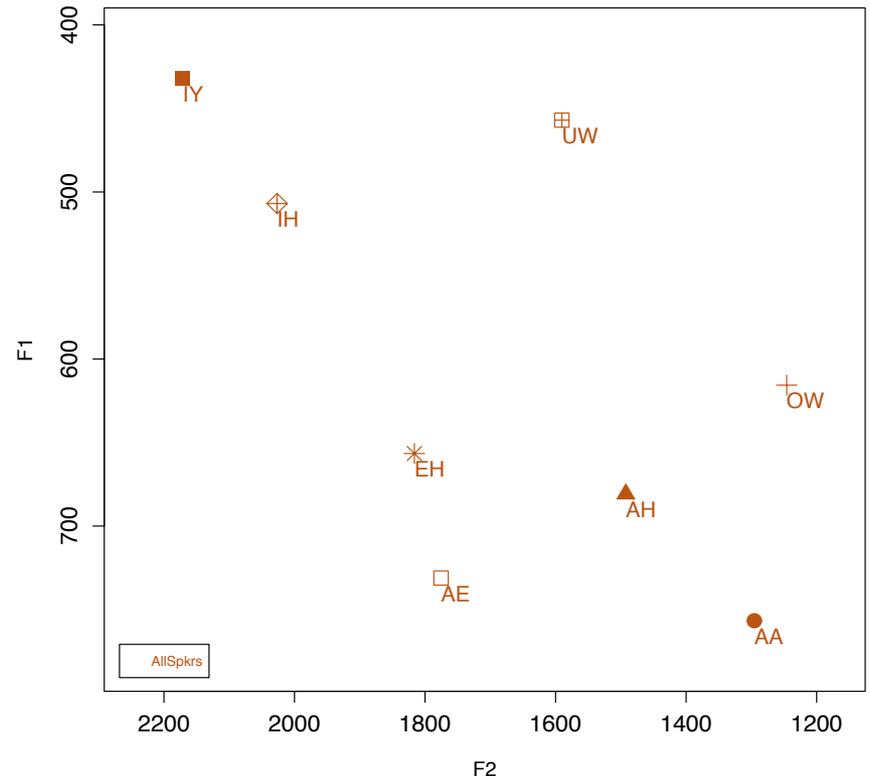
F2'

Toronto Anglo ENG vs CAN ENG

Toronto Anglo English



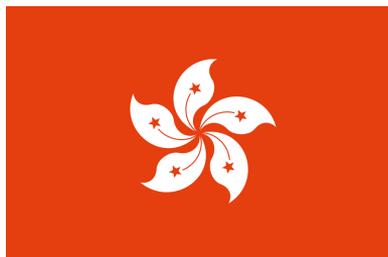
Toronto CAN Heritage English



Based on means from Roeder 2012,
Boberg 2008, Roeder & Jarmasz 2010

Based on means of 7 speakers

F1 and F2 Means for /i/ in open syllables



1st

2nd



Cantonese
(8 speakers)

Gen	F1*	F2*	Tokens
1	439	2044	3207
2	423	2106	857
All	435	2057	4064

- Gen 2 has higher and more fronted /i/
- *p < 0.05

CAN English
(11 speakers)

Gen	F1**	F2**	Tokens
1	454	2096	1545
2	434	2324	2370
All	441	2234	3925

- Gen 2 has higher and more fronted /i/
- **p < 0.01

Toronto Anglo English

F1	F2
474	2011

- Anglo English has the lowest /i/.

Discussion of Results

- Evidence of generational change clear with same general developmental trend in both languages.
 - Raising and fronting of /i/ for Gen 2 in both **CAN** and **CAN ENG**
- Relative position of /i/ and /ɪ/ are different in **CAN** and **ENG**.
- Lack of /u/ fronting in **CAN** observed, but some fronting in **CAN ENG**
- How these changes result from contact with English (if that is the case) appear to be quite complex
 - further research required to better understand how.
- Note
 - Tone not considered as a factor
 - Variation and change in other vowels not considered
 - No homeland data available

Discussion of Methodology

- Without human intervention, automatically extracted data creates reasonable vowel plots
- A promising avenue for future research on vowel variation and change in heritage languages
- But need to check and compare results with manual formant extraction

Future Work

- Assessing accuracy of automated alignment and formant extraction by attempting to replicate results using manual methods
- Expanding to more vowels and more speakers
 - 8 speakers for this analysis, ~ 40 **CAN** speakers in Corpus
 - Comparing homeland data
- Expanding to other heritage languages
 - Italian, Faetar, Russian, Ukrainian, Korean

HLVC RAs:

Cameron Abma

Vanessa Bertone

Ulyana Bila

Rosanna Calla

Minji Cha

Karen Chan

Joanna Chociej

Sheila Chung

Tiffany Chung

Courtney Clinton

Radu Craioveanu

Marco Covi

Derek Denis

Tonia Djogovic

Joyce Fok

Paolo Frasca

Matt Gardner

Rick Grimm

Dongkeun Han

Natalia Harhaj

Taisa Hewka

Melania Hrycyna

Michael Iannozzi

Diana Kim

Janyce Kim

Iryna Kulyk

Mariana Kuzela

Ann Kwon

Alex La Gamba

Carmela La Rosa

Natalia Lapinskaya

Kris Lee

Nikki Lee

Olga Levitski

Arash Lotfi

Paulina Lyskawa

Rosa Mastri

Timea Molnár

Jamie Oh

Maria Parascandolo

Rita Pang

Andrew Peters

Tiina Rebane

Hoyeon Rim

Will Sawkiw

Maksym Shkvorets

Vera Richetti Smith

Anna Shalaginova

Konstantin Shapoval

Yi Qing Sim

Mario So Gao

Awet Tekeste

Josephine Tong

Sarah Truong

Dylan Uscher

Ka-man Wong

Olivia Yu

Minyi Zhu

Collaborators:

Yoonjung Kang

Alexei Kochetov

James Walker

Funding:

SSHRC, University of

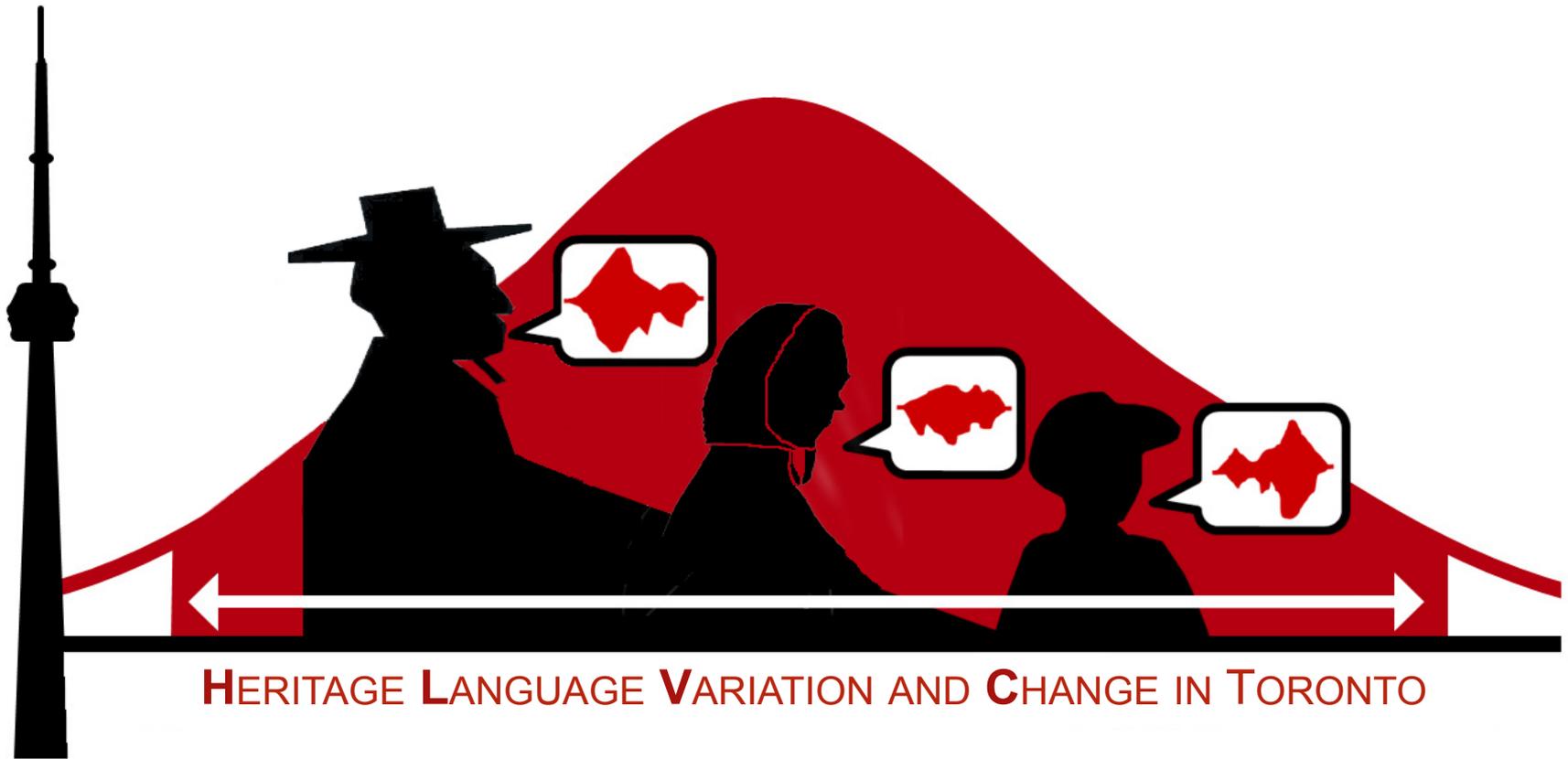
Toronto,

Shevchenko

Foundation

References

- Boberg, Charles. 2008. "Regional phonetic differentiation in Standard Canadian English." *Journal of English Linguistics* 36/2: 129-154.
- Gorman, Kyle, Jonathan Howell & Michael Wagner. (2011). Prosodylab-Aligner: A tool for forced alignment of laboratory speech. *Proceedings of Acoustics Week in Canada, Quebec City.*
- Hoffman, M. F., & Walker, J. A. (2010). Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *Language Variation and Change*, 22, 37-67.
- Lobanov
- Nagy, Naomi. (2009). *Heritage Language Variation and Change in Toronto*. <http://projects.chass.utoronto.ca/ngn/HLVC>.
- Roeder, Rebecca. 2012. "The Canadian Shift in Two Ontario Cities." *Special Issue of World Englishes: Autonomy and Homogeneity in Canadian English* 31,4: 478-492. Guest editors Stefan Dollinger and Sandra Clarke.
- Roeder, Rebecca and Lidia-Gabriela Jarmasz. 2010. "The Canadian Shift in Toronto." *Revue canadienne de linguistique/Canadian Journal of Linguistics* 55,3: 387-404.
- Rosenfelder, Ingrid; Fruehwald, Joe; Evanini, Keelan and Jiahong Yuan. (2011). FAVE (Forced Alignment and Vowel Extraction) Program Suite. <http://fave.ling.upenn.edu>.
- Wittenburg, Peter, H. Brugman, Albert Russel, A. Klassmann, and Han Sloetjes. (2006). *ELAN: a Professional Framework for Multimodality Research*. *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*.
- Yue Hashimoto, Oi-kan 1972. *Phonology of Cantonese*. Cambridge University Press.



HERITAGE LANGUAGE VARIATION AND CHANGE IN TORONTO

[HTTP://PROJECTS.CHASS.UTORONTO.CA/NGN/HLVC](http://projects.chass.utoronto.ca/ngn/hlvc)

EMAIL: NAOMI.NAGY@UTORONTO.CA

FOR TODAY'S SLIDES: EMAIL: HBT3@PITT.EDU