OzCLO

First Australian Computational and Linguistics Olympiad

NATIONAL ROUND

Problem set
(2 hours)

YOU ARE ONLY ALLOWED TO WORK ON THE PROBLEM SET IN TWO HOURS.
RETURN IT TO THE FACILITATOR AT THE END OF THE CONTEST TIME.

(A)Íslenska [Icelandic]

Íslenska is the national language of Iceland, and is spoken by around 230,000 people in Iceland and more than 5,000 people in Canada.

1a. Hin fagra lilja er gul.	The beautiful lily is yellow.
b. Fögur lilja er gul.	A beautiful lily is yellow.
c. Hin gula lilja er fögur.	The yellow lily is beautiful.
d. Gul lilja er fögur.	A yellow lily is beautiful.
e. Liljan er fögur.	The lily is beautiful.
2a. Hinn guli akur er fagur.	The yellow field is beautiful.
b. Akurinn er gulur.	The field is yellow.
c. Gulur akur er fagur.	A yellow field is beautiful.
d. Hinn fagri akur er gulur.	The beautiful field is yellow.
3a. Hith fagra hreithur er gult.	The beautiful nest is yellow.
b. Fagurt hreithur er gult.	A beautiful nest is yellow.
c. Hreithrith er gult.	The nest is yellow.
d. Hith gula hreithur er fagurt.	The yellow nest is beautiful.
4a. Nyrath er gult.	The kidney is yellow.
b. Hith fagra nyra er gult.	The beautiful kidney is yellow.
c. Hinn fagri hestur er gulur.	The beautiful horse is yellow.
d. Fagur hestur er gulur.	A beautiful horse is yellow.
e. Lifrin er fögur.	The liver is beautiful.
f. Gul lifur er fögur.	A yellow liver is beautiful.

A1.   Translate the following sentences into Íslenska.

(a) The horse is beautiful.

(b) A yellow nest is beautiful.

(c) The yellow kidney is beautiful.

B2. The following sentences are ungrammatical.   Write the grammatical translations.

(a) *Hin lilja er gul   The lily is yellow.

_____

(b) *Fagur lifur er gulur.   A beautiful liver is yellow.

_____

(c) *Hestrinn er gulur.   The horse is yellow.

_____

B3. Explain briefly for each ungrammatical sentence in 2. why it is ungrammatical. Refer to the numbers of the example sentences which support your claims.

Problem created by Jane Simpson 2008

(B) Manam, I'm Anam

Manam Pile ("Manam Talk") is a Malayo-Polynesian language spoken on Manam Island off the coast of Papua New Guinea. Manam is one of the most active volcanoes in the world, and during violent eruptions the population must be evacuated to the mainland.
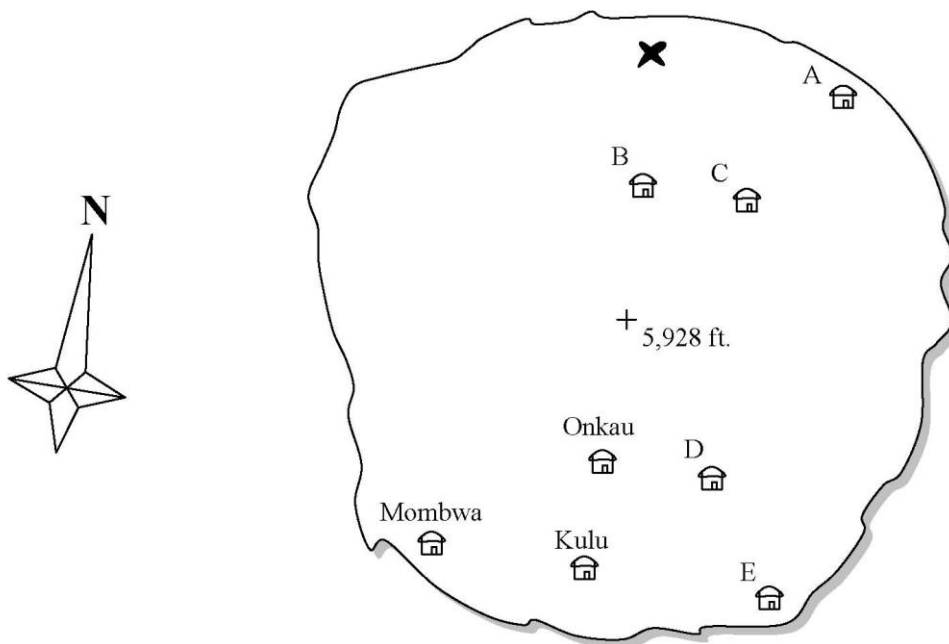
Below, a Manam islander describes the relative locations of the houses above.

1   *Onkau pera kana auta ieno, Kulu pera kana ilau ieno.*
2   *Mombwa pera kana ata ieno, Kulu pera kana awa ieno.*
3   *Tola pera kana auta ieno, Sala pera kana ilau ieno.*
4   *Sulung pera kana awa ieno, Tola pera kana ata ieno.*
5   *Sala pera kana awa ieno, Mombwa pera kana ata ieno.*
6   *Pita pera kana ilau ieno, Sulung pera kana auta ieno.*
7   *Sala pera kana awa ilau ieno, Onkau pera kana ata auta ieno.*
8   *Butokang pera kana awa auta ieno, Pita pera kana ata ilau ieno.*

B1. Onkau's, Mombwa's, and Kulu's houses have already been located on the map above. Who lives in the other five houses?

A: _____ B: _____ C: _____ D: _____ E: _____

B2. Arongo is building his new house in the location marked with an X. In three Manam Pile sentences like the ones on the previous page, describe the location of Arongo's house in relation to the three closest houses.

1 _____

2 _____

3 _____

B3. Explain your answers to B1 and B2 in the space provided below.

(C) aw-TOM-uh-tuh

Finite-state automata (FSA) are a type of abstract "machine" with many possible uses. One possible use is to guess what language a document (such as a webpage) is in. If we make an automaton that can distinguish between possible English words and impossible ones, and then give it a webpage with a bunch of words that are impossible in English (like "*aioaepa*" or "*ragaiiare*"), we can be pretty sure that the webpage isn't written in English. (Or, at least, isn't *entirely* written in English.)

Here is a finite state automaton that can distinguish between possible and impossible words in Rotokas, a language spoken on the island of Bougainville off the coast of New Guinea. Rotokas has a very simple system of sounds and allows us to create a very small FSA.
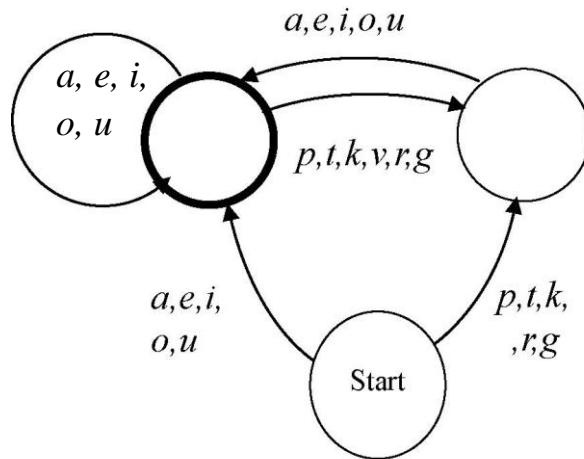
An FSA works like a board game. Choose a word, and place your pencil on the space marked "Start". Going through the letters of the word one at a time, move your pencil along the path marked with that letter. If the word ends and you're at a space marked with a thicker circle, the word succeeds: it's a possible Rotokas word! If the word ends and you're not at a thicker circle, or you're midway through the word and there's no path corresponding to the next letter, the word fails: it's not a possible Rotokas word!

Try it out with these possible and impossible words; the automaton should accept all the possible words and reject the impossible ones.

Possible Rotokas words		Impossible Rotokas words	
tauo	*kareveiepa*	*grio*	*ouag*
puraveva	*ovokirovuia*	*ovgi*	*vonoka*
avaopa	*ouragaveva*	*gataap*	*oappa*

C1. Now, using the automaton above, put a check mark next to each possible Rotokas word:

___ *iu*	___ *uente*	___ *voav*
___ *idau*	___ *urioo*	___ *uaia*
___ *oire*	___ *raorao*	___ *oratreopaveiepa*

C2. Actually, the system that people on Bougainville use to write Rotokas is a little more complicated than the one we've presented here. In addition to the eleven letters above, the real Rotokas alphabet has a twelfth letter, S. This letter represents the sound "s" or "ts", which in Rotokas only occurs in very specific situations.

Below is a skeleton of an FSA for Rotokas with all the path labels removed and set to one side.

a,e,i,o,u	*i*
t	*a,e,i,o,u*
a,e,o,u	*s*
p,k,v,r,g	
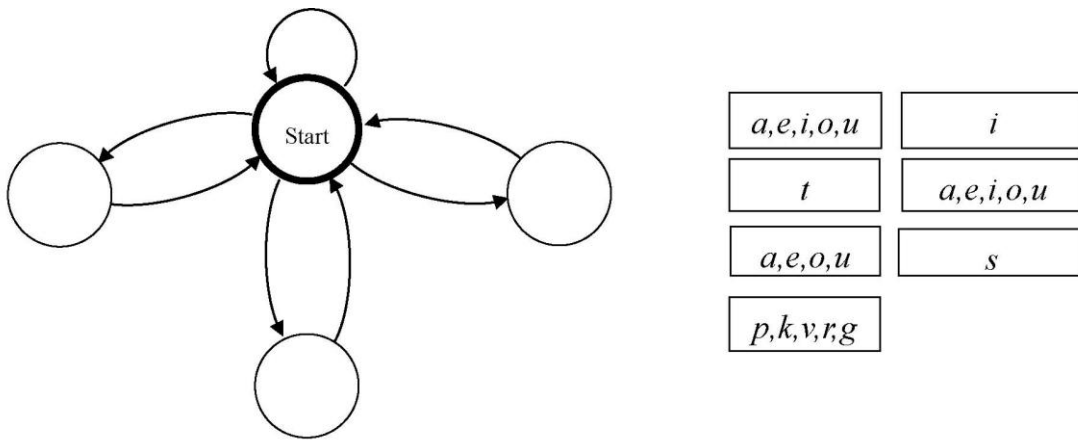
Replace the path labels so that exactly half of the words below succeed and the other half fail.

oisio	*tiravau*	*saiuu*	*kotoe*
uasau	*utsa*	*sioparoia*	*parauos*
puapuata	*sisigarue*	*porouativeve*	*aasiia*

C3. Why do T and S get their own paths? What is special about these letters?

(D)Tzolk'in

Central to Mayan thought and religion was the concept of time, which was held to be cyclical and without beginning or end. Accurate reckoning of the day was highly important for ceremonial, political, and agricultural reasons, and to this end the Mayas kept several calendars of different length, all cycling simultaneously, while also keeping track of the exact number of days since August 11, 3114 BC.
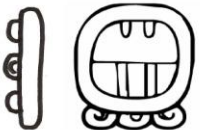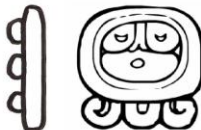
The *tzolk'in* ritual and ceremonial calendar is the oldest and most important of these, and is not exclusively Mayan – it was shared by all of the ancient civilizations of Mexico, and predates Classical Mayan civilization by hundreds of years. It is also the main calendar to have survived, and is still in use today in some Mayan villages.

Given the conception of time as cyclical, it is appropriate that the *tzolk'in* has no identifiable beginning or end, but simply assigns names to days in an infinite loop. Below we've taken a calendar for our August and September, but inserted the day names according to the *tzolk'in*.

D1. What are the Mayan names of the days labeled *a* and *b* on the calendar? Draw the appropriate glyphs in the boxes below.

D2. Where on the calendar would the following days fall? Write a *c* and a *d* on the appropriate days on the following page.

D3. The two months shown are completely representative of the workings of the *tzolk'in*; there are no special sub-cycles, "leap" days, periods with fewer days than other periods, etc. Knowing that, *how often* does the following day occur?

AUGUST

SEPTEMBER

b

a

D4. Explain your answers in the space provided below.

Patrick Littell and Erin Donnelly

(E)The Whole Spectrum

Background

The sounds of human speech (and indeed all sounds) travel through the air in waves, some of which your ear detects as sounds. These waves can be analyzed (by a mathematical technique called "Fourier analysis") into combinations of basic ("sinusoidal") waves whose most important properties are "frequency" and "amplitude." One sound may comprise many such basic waves, and a basic wave of the same frequency or amplitude may appear in many sounds. Linguists sometimes display this analysis on a diagram called a spectrogram: the resulting sine waves' frequencies are plotted vertically (with greater amplitude indicated by darker points) and the time horizontally.

Data

A sequence of spectrograms produced using the computer software "Praat" is shown in the following pages. The first 12 are each labeled with an English word, which is shown in the spectrogram. The following words are given:

sash, lamb, knee, sheesh, soup, pang, loose, (the letter) e, mice, ice, coo, shine

The last four spectrograms each show one of the following eight words:

louse, lass, lease, lice, pass, ash, sheep, lack

Problems

E1. What words are shown in the last four spectrograms? You may give up to two answers on each; if you give two, you will be given half credit if either of them is correct.

(13) _____   (14) _____   (15) _____   (16) _____

E2. For each of the three most significant sounds in "sash," mark an interval corresponding to the duration of that sound. You should mark each interval by a horizontal line drawn above the spectrogram over all parts of the spectrogram affected by that sound. If you make other markings on the spectrogram, clearly indicate where each of the three requested intervals starts and ends.

E3. Do the same for "lamb."

E4. Explain your reasoning for E1.

E5. Discuss the correspondence between the English spellings of the given 12 words and their spectrograms. In particular, some letters' sounds affect longer intervals in the spectrograms than others do. Which letters and sounds affect the longest intervals? Do adjacent sounds' affected intervals overlap, and if so, how? Are there any letters that correspond to no interval at all, and which letters? Conversely, are there any distinct sounds on the spectrograms that correspond to no English letters, and where are those sounds?
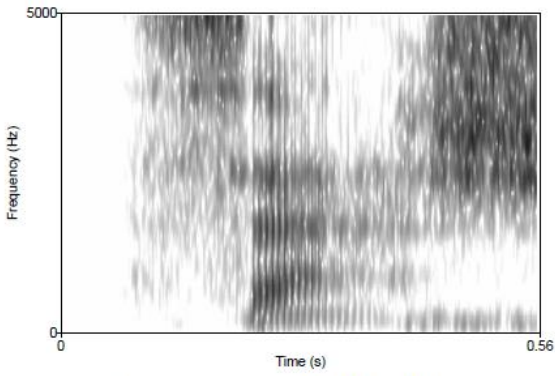
Spectrogram 1: "Sash"
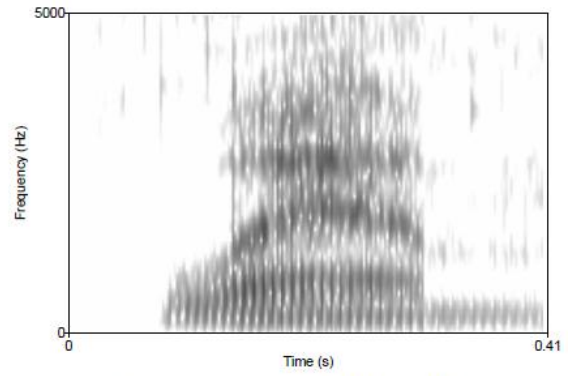
Spectrogram 2: "Lamb"
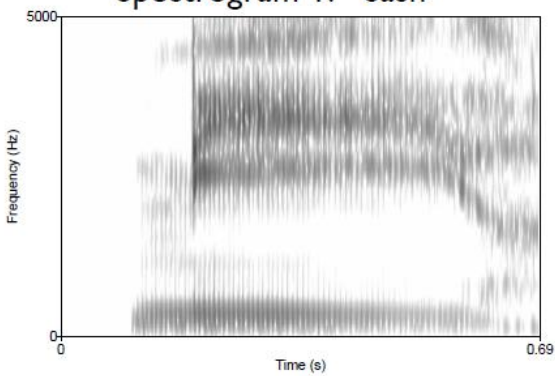
Spectrogram 3: "Knee"
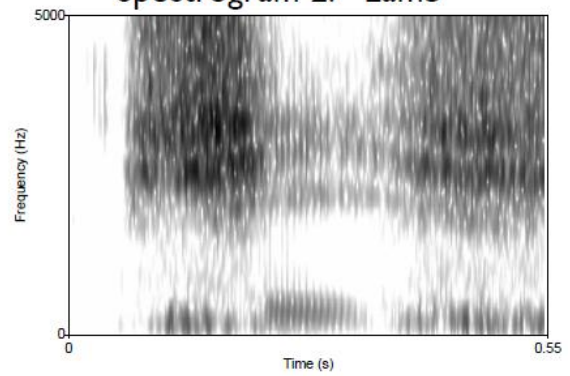
Spectrogram 4: "Sheesh"
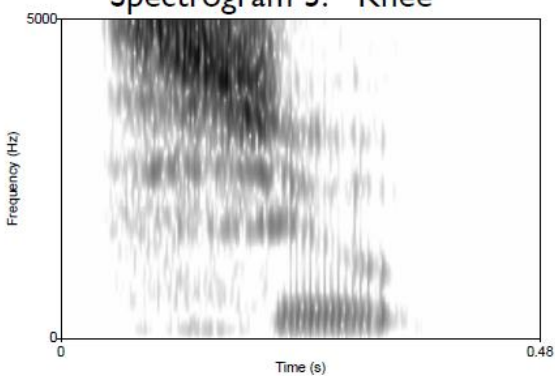
Spectrogram 5: "Soup"
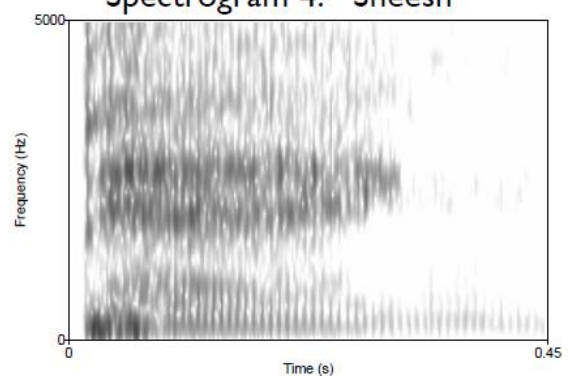
Spectrogram 6: "Pang"
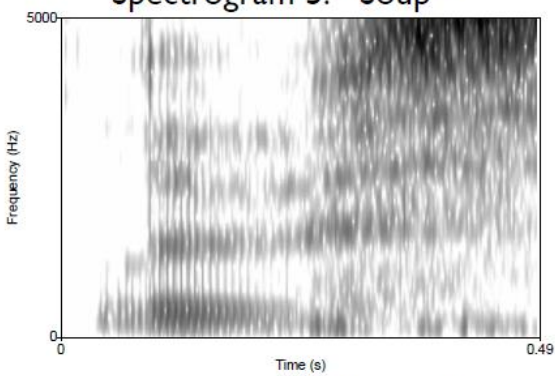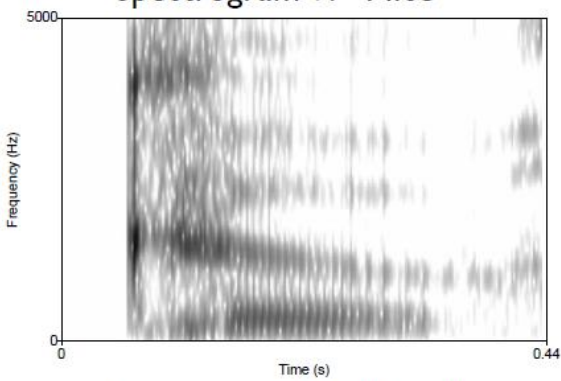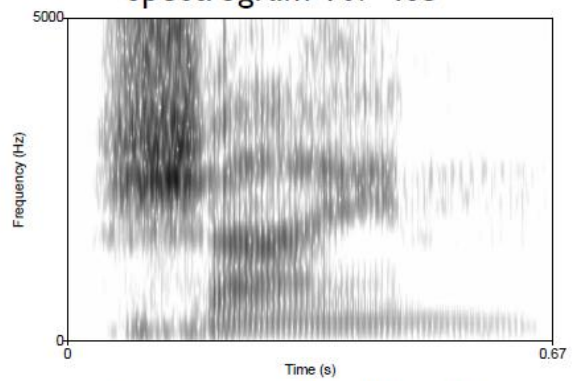

Spectrogram 7: "Loose"
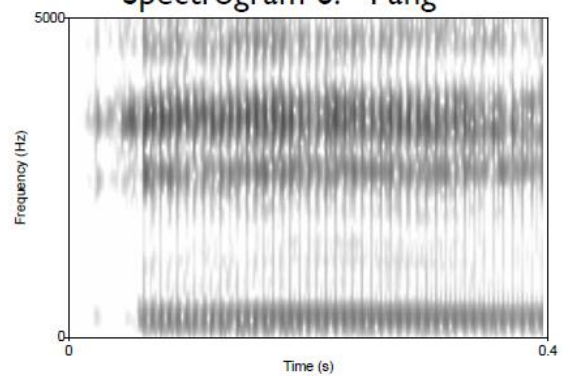
Spectrogram 8: (the letter) `e'
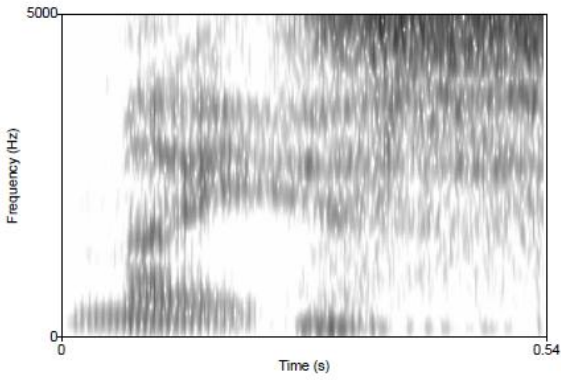
Spectrogram 9: "Mice"
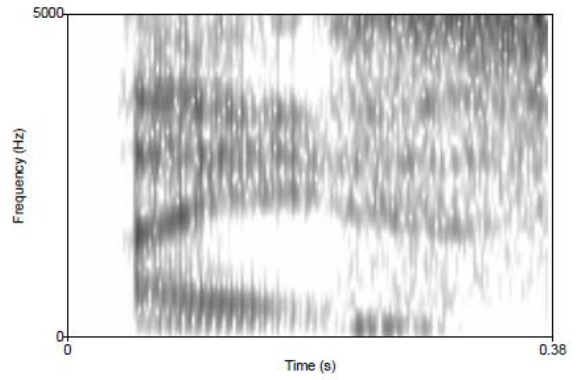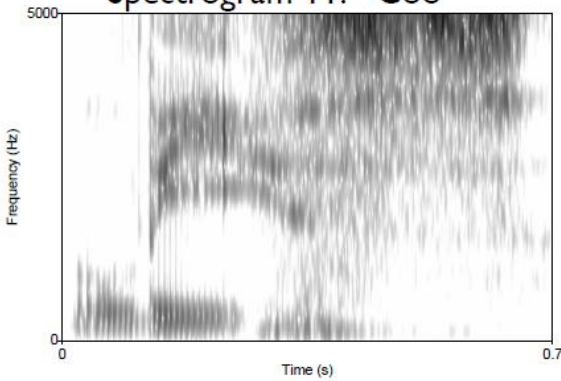

Spectrogram 10: "Ice"
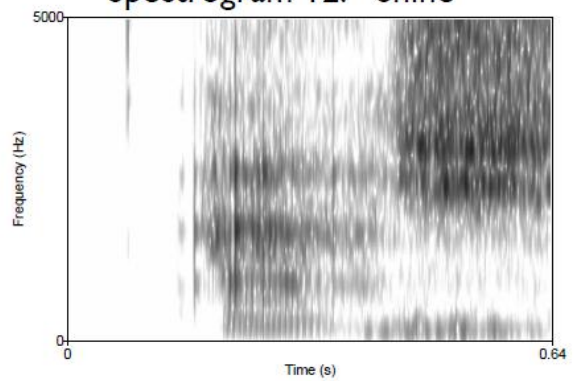

Spectrogram 11: "Coo"
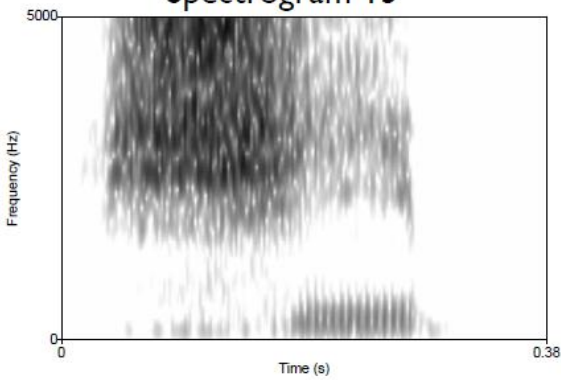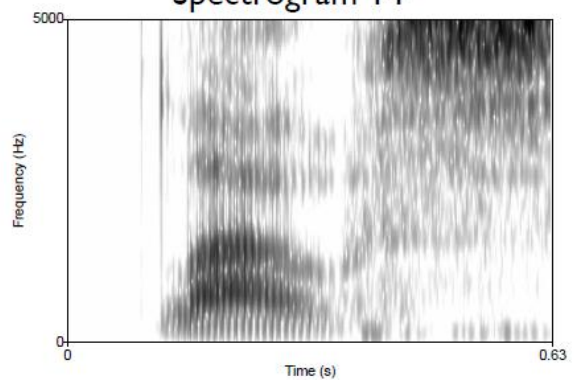

Spectrogram 12: "Shine"

Spectrogram 13

Spectrogram 14

Spectrogram 15

Spectrogram 16

Problem created by Adam Hesterberg