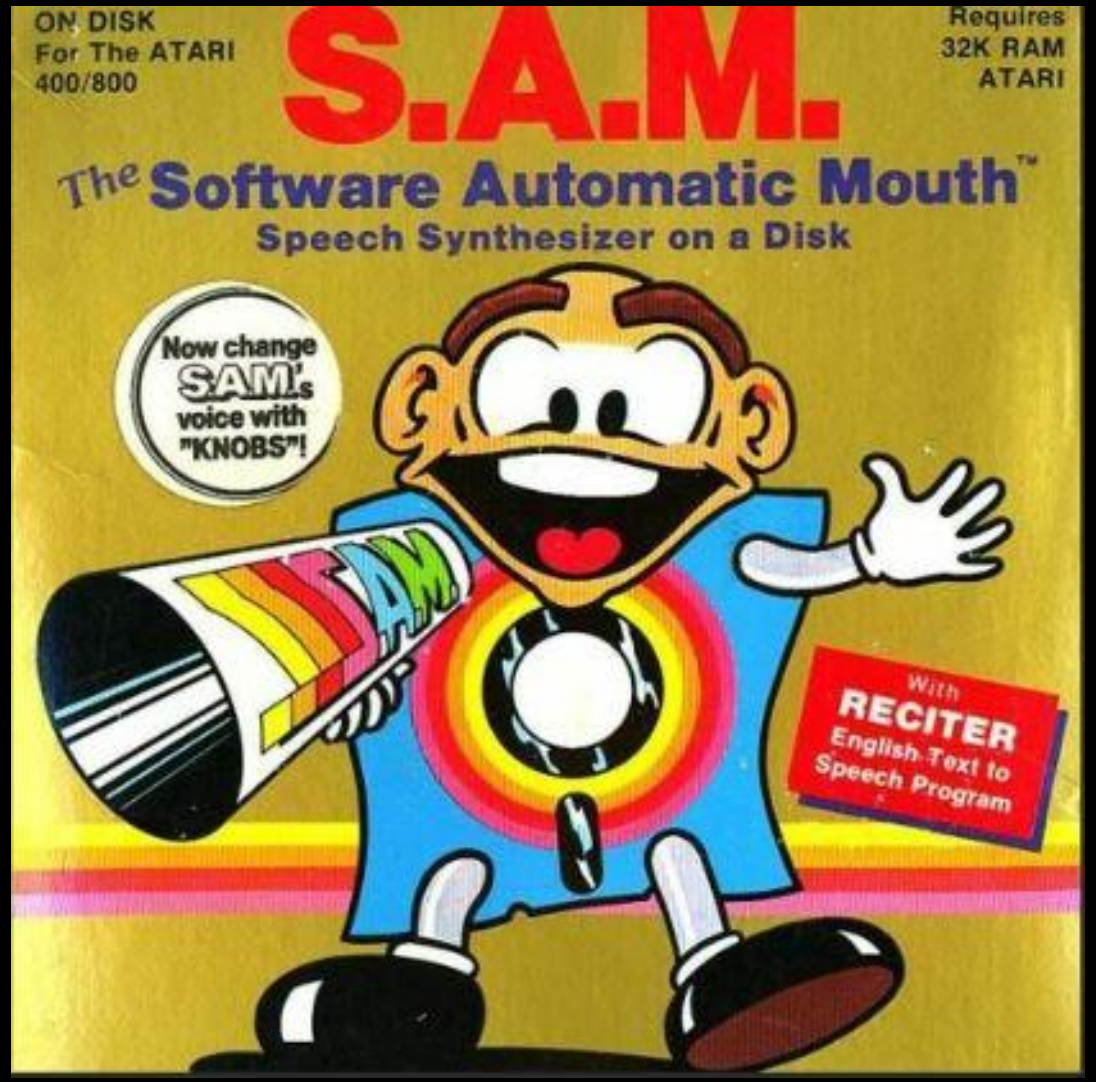
Text-To-Speech From Scratch

Dylan Oonk


ON DISK
For The ATARI
400/800
S.A.M.
Requires
32K RAM
ATARI
The Software Automatic Mouth™
Speech Synthesizer on a Disk
Now change SAM's voice with "KNOBS"!
SAM
With
RECITER
English Text to Speech Program

Normalization

123 → One hundred twenty three

Tokenization

“Hello, World!” → 1. “Hello” 2. “World”

Phoneme…ization

“Dallas” → [ˈdæləs]

Synthesization

[əˈmɛrɪkə] → 

Normalization

```
# Titles
(re.compile(r'\bmr\.'), 'mister'),
(re.compile(r'\bmrs\.'), 'missus'),
(re.compile(r'\bms\.'), 'miss'),
(re.compile(r'\bdr\.'), 'doctor'),
(re.compile(r'\bprof\.'), 'professor'),
(re.compile(r'\brev\.'), 'reverend'),
(re.compile(r'\bgen\.'), 'general'),
(re.compile(r'\bsen\.'), 'senator'),
(re.compile(r'\brep\.'), 'representative'
(re.compile(r'\bgov\.'), 'governor'),
(re.compile(r'\bcol\.'), 'colonel'),
(re.compile(r'\bcapt\.'), 'captain'),

# Common abbreviations
(re.compile(r'\be\.g\.'), 'for example'),
(re.compile(r'\bi\.e\.'), 'that is'),
(re.compile(r'\betc\.'), 'et cetera'),
(re.compile(r'\bvs\.'), 'versus'),
(re.compile(r'\bvs\b'), 'versus'),
(re.compile(r'\bet al\.'), 'and others'),
(re.compile(r'\bapprox\.'), 'approximatel
(re.compile(r'\bdept\.'), 'department'),
(re.compile(r'\bfig\.'), 'figure'),
(re.compile(r'\bno\.'), 'number'),
(re.compile(r'\bpg\.'), 'page'),
(re.compile(r'\bvol\.'), 'volume'),
(re.compile(r'\bch\.'), 'chapter'),
(re.compile(r'\bsec\.'), 'section'),
```

```
# Time
(re.compile(r'\b(a\.m\.)'), ' a m '),
(re.compile(r'\b(p\.m\.)'), ' p m '),

# Units (avoiding numbers)
(re.compile(r'(?<=\d)\s*ft\.'), ' feet '),
(re.compile(r'(?<=\d)\s*in\.'), ' inches '),
(re.compile(r'(?<=\d)\s*lb\.'), ' pounds '),
(re.compile(r'(?<=\d)\s*lbs\.'), ' pounds '),
(re.compile(r'(?<=\d)\s*oz\.'), ' ounces '),

# Organizations
(re.compile(r'\bcorp\.'), 'corporation'),
(re.compile(r'\bco\.'), 'company'),
(re.compile(r'\binc\.'), 'incorporated'),
(re.compile(r'\bltd\.'), 'limited'),
(re.compile(r'\bl\.l\.c\.'), ' l l c '),

# Directions
(re.compile(r'\bn\.'), 'north'),
(re.compile(r'\bs\.'), 'south'),
(re.compile(r'\be\.'), 'east'),
(re.compile(r'\bw\.'), 'west'),
(re.compile(r'\bne\b'), 'north east'),
(re.compile(r'\bnw\b'), 'north west'),
(re.compile(r'\bse\b'), 'south east'),
(re.compile(r'\bsw\b'), 'south west'),
```

```
# Parentheses and brackets
(re.compile(r'\('), ' open paren
(re.compile(r'\)'), ' close pare


(re.compile(r'"'), ' [QUOTE] '),

# Slashes
(re.compile(r'/'), ' slash '),
(re.compile(r'\\'), ' back slash

# Ampersand
(re.compile(r'&'), ' and '),

# At symbol
(re.compile(r'@'), ' at '),

# Hashtag
(re.compile(r'#'), ' hashtag '),

# Percent sign
(re.compile(r'%'), ' percent '),

# Asterisk
(re.compile(r'\*'), ' asterisk '
```

```
# Ampersand
(re.compile(r'&'), ' and '),
```

Pattern

Replacement

\bdr\.

doctor

“i saw doctor smith”

```
--- Running replace_punctuation_and_expand_abreviations tests ---
All Tests Passed For replace_punctuation_and_expand_abreviations()

--- Running numbers_to_words tests ---
All Tests Passed For numbers_to_words()

--- Running normalize_text tests ---
All Tests Passed For normalize_text()
```

Tokenization

The `split()` method splits a string into a list.

```
token = Token("Hello")
tokens = TokenList(["hello", "world"])
```

Phonemicization

Ghoti

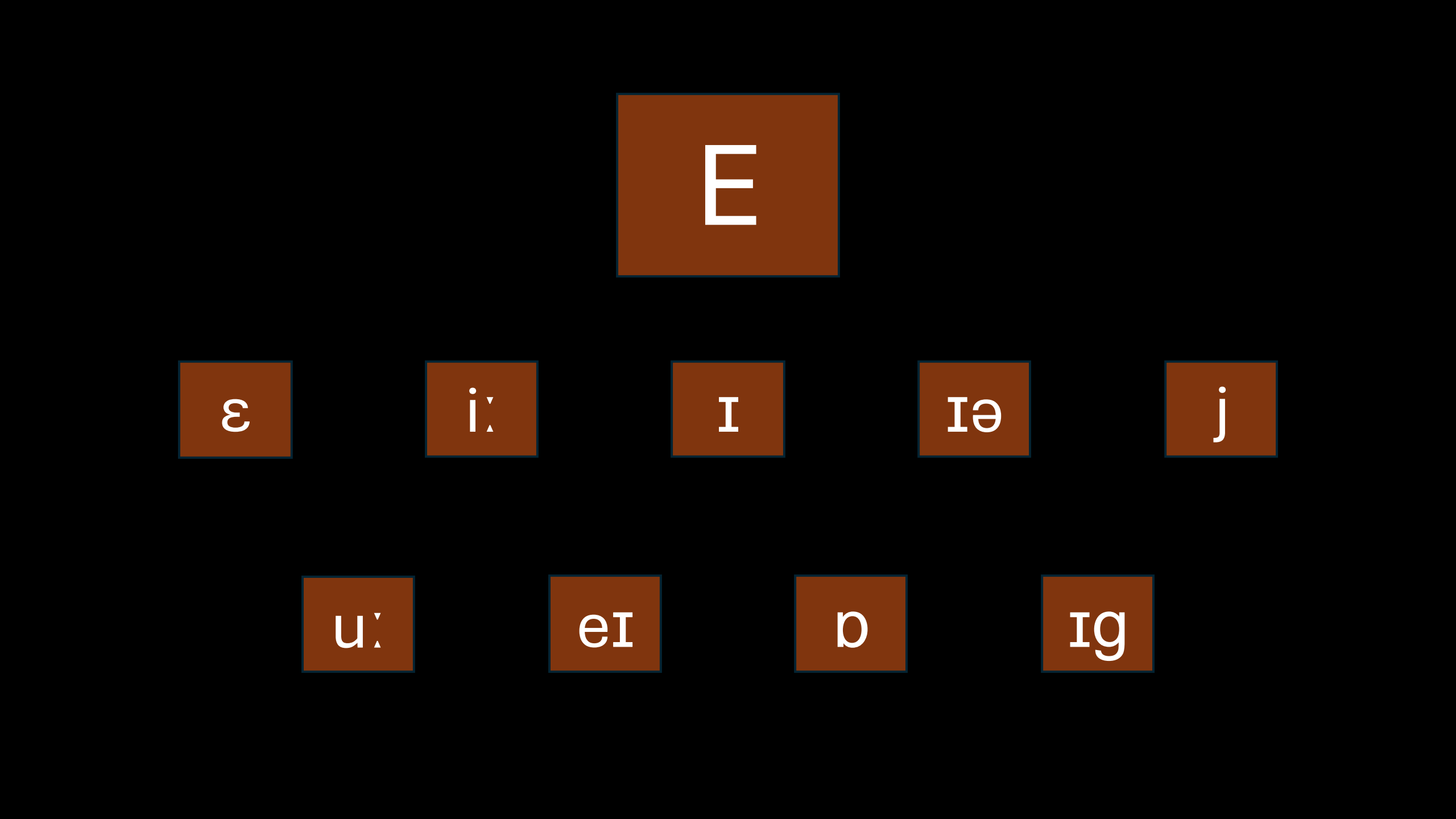
27 languages

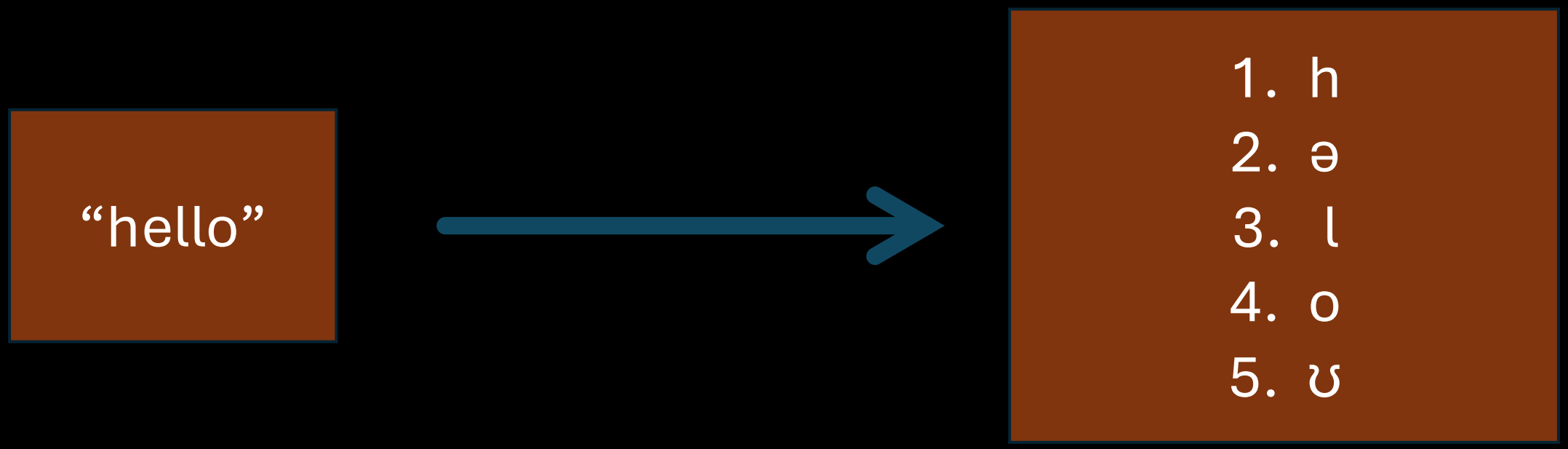
Article Talk Read Edit View history Tools

From Wikipedia, the free encyclopedia

> *For other uses, see Ghoti (disambiguation).*

Ghoti is a creative English respelling of the word *fish*, used to illustrate irregularities in English spelling and pronunciation.


Ɛ
ɛ
iː
ɪ
ɪə
j
uː
eɪ
ɒ
ɪg

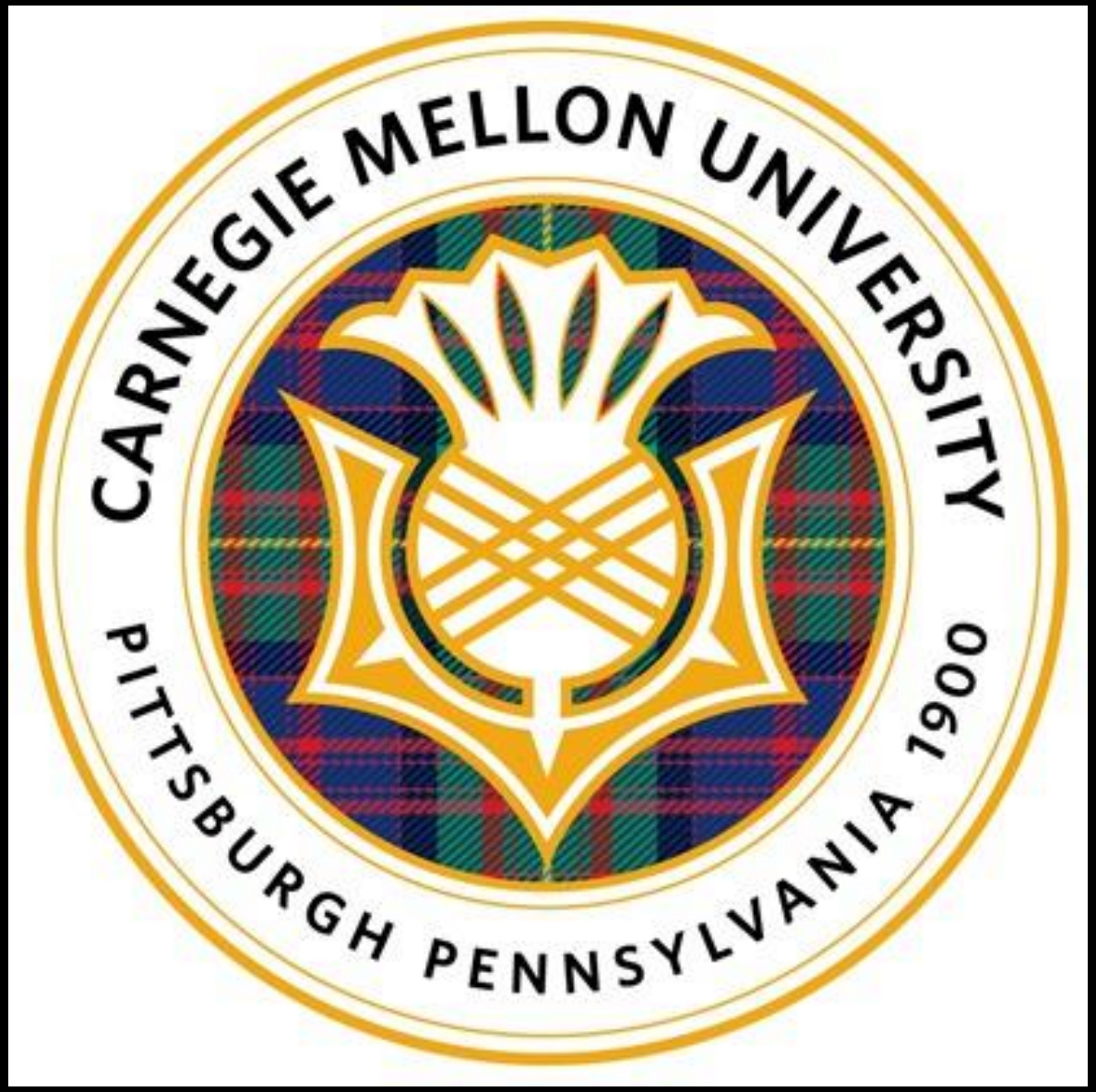

“hello”
1. h
2. ə
3. l
4. o
5. ʊ

ARPA

ARPABET		IPA	Example(s)
1-letter	2-letter		
a	AA	ɑ~ɒ	balm, bot (with father–bother merger)
@	AE	æ	bat
A	AH	ʌ	butt
c	AO	ɔ	caught, story
W	AW	aʊ	bout
x	AX	ə	comma
—	AXR[3]	ɚ	letter, forward
Y	AY	aɪ	bite
E	EH	ɛ	bet
R	ER	ɝ	bird, foreword
e	EY	eɪ	bait
I	IH	ɪ	bit
X	IX	ɨ	roses, rabbit
i	IY	i	beat
o	OW	oʊ	boat
O	OY	ɔɪ	boy
U	UH	ʊ	book
u	UW	u	boot
—	UX[3]	ʉ	dude

ARPABET		IPA	Example
1-letter	2-letter		
b	B	b	buy
C	CH	tʃ	China
d	D	d	die
D	DH	ð	thy
F	DX	ɾ	butter
L	EL	l̩	bottle
M	EM	m̩	rhythm
N	EN	n̩	button
f	F	f	fight
g	G	g	guy
h	HH *or* H[3]	h	high
J	JH	dʒ	jive
k	K	k	kite
l	L	l	lie
m	M	m	my
n	N	n	nigh
G	NX *or* NG[3]	ŋ	sing
—	NX[3]	ɾ̃	winter
p	P	p	pie
Q	Q	ʔ	uh-oh
r	R	ɹ	rye
s	S	s	sigh
S	SH	ʃ	shy
t	T	t	tie
T	TH	θ	thigh
v	V	v	vie
w	W	w	wise
H	WH	ʍ	why (without wine–whine merger)
y	Y	j	yacht
z	Z	z	zoo
Z	ZH	ʒ	pleasure

```
dalke D EY1 L K
dalkon D AE1 L K AH0 N
dall D AO1 L
dalla D AE1 L AH0
dallaire D AA1 L EH0 R
dallara D AE2 L AA1 R AH0
dallas D AE1 L AH0 S
dallas' D AE1 L AH0 S
dallas's D AE1 L AH0 S IH0 Z
dalley D AE1 L IY0
dallhold D AO1 L HH OW2 L D
dalliance D AE1 L IY0 AH0 N S
```

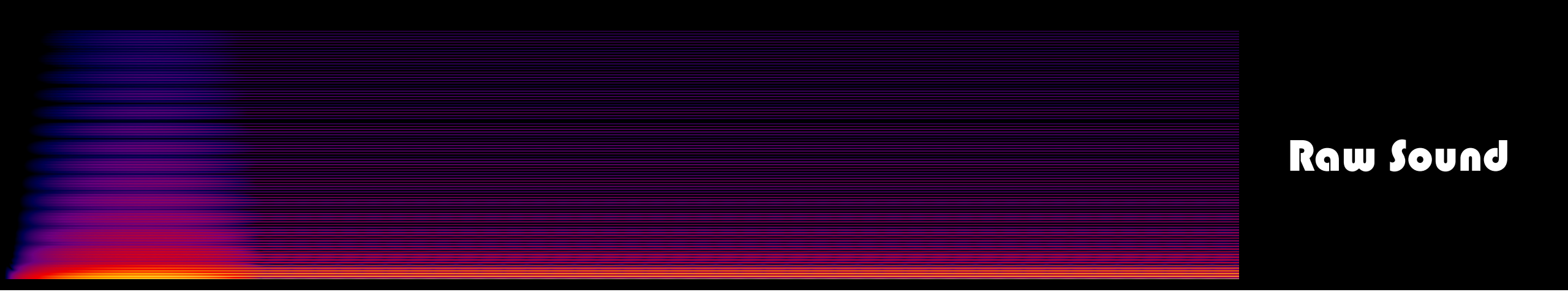

CARNEGIE MELLON UNIVERSITY
PITTSBURGH PENNSYLVANIA 1900

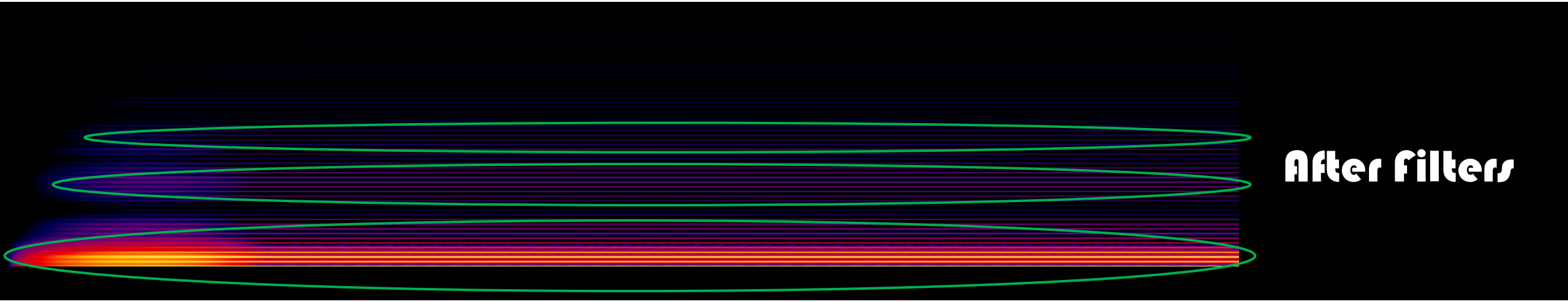
Rows: 126,052

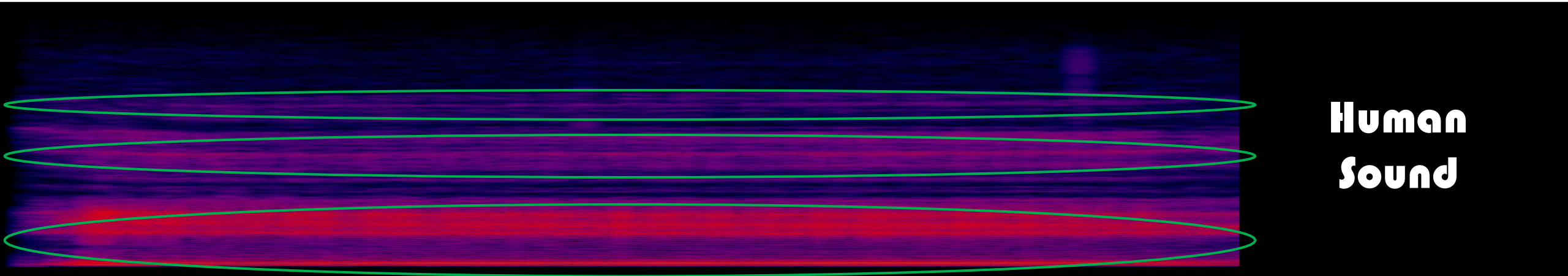
	word	sound
	Filte	Filte
1	'bout	B AW1 T
2	'cause	K AH0 Z
3	'course	K AO1 R S
4	'cuse	K Y UW1 Z
5	'em	AH0 M
6	'frisco	F R IH1 S K OW0

```python
def fallback_pronunciation(TEXT: str, CONNECTION: sqlite3.Connection, CURS
    CHARACTERS: list[str] = list(TEXT.strip())
    output: list[str] = []

    for CHARACTER in CHARACTERS:
        sound = get_sound_from_db(CHARACTER.lower(), CONNECTION, CURSOR)
        output.append(sound)

    return " ".join(output).strip()
```

Synthesization


Raw Sound


After filters


Human
Sound

Control Methods Used in a Study of the Vowels

GORDON E. PETERSON AND HAROLD L. BARNEY
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey

(Received December 3, 1951)

Relationships between a listener's identification of a spoken vowel and its properties as revealed from acoustic measurement of its sound wave have been a subject of study by many investigators. Both the utterance and the identification of a vowel depend upon the language and dialectal backgrounds and the vocal and auditory characteristics of the individuals concerned. The purpose of this paper is to discuss some of the control methods that have been used in the evaluation of these effects in a vowel study program at Bell Telephone Laboratories. The plan of the study, calibration of recording and measuring equipment, and methods for checking the performance of both speakers and listeners are described. The methods are illustrated from results of tests involving some 76 speakers and 70 listeners.

TABLE II. Averages of fundamental and formant frequencies and formant amplitudes of vowels by 76 speakers.

		i	ɪ	ɛ	æ	ɑ	ɔ	ʊ	u	ʌ	ɝ
Fundamental frequencies (cps)	*M*	136	135	130	127	124	129	137	141	130	133
	W	235	232	223	210	212	216	232	231	221	218
	Ch	272	269	260	251	256	263	276	274	261	261
Formant frequencies (cps)											
F_1	*M*	270	390	530	660	730	570	440	300	640	490
	W	310	430	610	860	850	590	470	370	760	500
	Ch	370	530	690	1010	1030	680	560	430	850	560
F_2	*M*	2290	1990	1840	1720	1090	840	1020	870	1190	1350
	W	2790	2480	2330	2050	1220	920	1160	950	1400	1640
	Ch	3200	2730	2610	2320	1370	1060	1410	1170	1590	1820
F_3	*M*	3010	2550	2480	2410	2440	2410	2240	2240	2390	1690
	W	3310	3070	2990	2850	2810	2710	2680	2670	2780	1960
	Ch	3730	3600	3570	3320	3170	3180	3310	3260	3360	2160
Formant amplitudes (db)	L_1	−4	−3	−2	−1	−1	0	−1	−3	−1	−5
	L_2	−24	−23	−17	−12	−5	−7	−12	−19	−10	−15
	L_3	−28	−27	−24	−22	−28	−34	−34	−43	−27	−20

Live Demo

(Don't mess it up)