

Can the Computer, and the Public, do the Metadata Work?

Karen Cariani

AAPB Project Director, WGBH
Executive Director, WGBH Media
Library & Archives

Casey Davis-Kaufman

AAPB Project Manager, WGBH
Associate Director, WGBH Media
Library & Archives

Agenda

- Who we are
- The project
 - Cataloguing dilemma
 - Audio cataloging
 - Speech to text results
 - Crowd source work
 - Crowd source challenges
- CL tool data output – useful?
- Additional available tools- work with CL experts
 - We have a great data set
 - You have expertise



WHO WE ARE



American Archive

OF PUBLIC BROADCASTING

American Archive of Public Broadcasting

The Library of Congress
Packard Campus for Audio Visual Conservation



Who we are:

WGBH Media Library and Archives



Discover historic programs of publicly funded radio and television across America. Watch and listen.

[Participating Organizations](#)

[Curated Exhibits](#)

[Browse The AAPB](#)

[Telling Our Story](#)

Search...



[Advanced Search...](#)



Interview with Presidential Candidate George Wallace
WGBH



Idaho Reports: Right to Work
Idaho Public Television



Conversations: International Museum of Muslim Cultures
Mississippi Public Broadcasting



60 years of recorded history, at your fingertips.

Watch and listen at americanarchive.org



50,000 HOURS

of historic public television and radio content are digitized and preserved



ACCESS

the entire collection at WGBH and the Library of Congress



WATCH & LISTEN

to more than 20,000 programs online at americanarchive.org



SEARCH

2.5 million additional catalog records in our database



DISCOVER

content from more than 100 organizations in 40 states and territories



EXPLORE

curated exhibits on topics such as civil rights, climate change, elections, and protesting

FIX IT

HELP

make the AAPB easier to search and access! Play our FIXIT game at fixit.americanarchive.org

Contact Information

aapb_notifications@wgbh.org



Facebook

facebook.com/amarchivpub



Twitter

[@amarchivpub](https://twitter.com/amarchivpub)

COLLABORATORS



LIBRARY OF CONGRESS

FUNDERS



Corporation for Public Broadcasting



Council on Library and Information Resources



Museum of Modern Art Library



National Endowment for the Humanities

THE PROJECT

American Archive

OF PUBLIC BROADCASTING



pop^{up} archive



The University of Texas at Austin
School of Information



INSTITUTE of
Museum and Library
SERVICES

the situation/dilemma

90,000 digitized television and radio programs

incomplete, inaccurate metadata records

limited staff resources

what is in the collection?

users need access to the collection

continued growth of the collection (content and sparse metadata)

Record

The MacNeil/Lehrer Report; Square Tomatoes; Tomatoes



Transcript

Search...

1 2 3 4

Hide

WOMAN IN SUPERMARKET: These are better than most that I've seen. Tomatoes in general have been so disappointing that I've gotten to the point where I just walk by them; I don't buy them any more because either they're very hard and when you eat them they have no taste at all. They're very waxy. Very often I've found them rotten inside and you can't seem to tell it from the outside. They are terribly expensive, and we love them because we're big salad eaters, you know. But they're not the kind of tomatoes we used to get years ago.

ROBERT MACNEIL: Good evening. Most of this week we're all going to be preoccupied with the energy crisis, but tonight we'd like to remind you that this country faces other crises than energy. And one of them is the tomato. Jim?

JIM LEHRER: It's a crisis of confidence, Robin. Recent opinion polls ranked the tomato among the least loved of all vegetables. Turned-off tomato eaters

If you have more information about this item than what is given here, we want to know! Contact us, indicating the AAPB ID (cpb-aacip/507-x31ng4hp5t).

Series	The MacNeil/Lehrer Report	Description	No description available This episode features a discussion on tomatoes. The guests are Karen Hess, Jim Hightower, Wayne Hawkins. Byline: Robert MacNeil, Jim Lehrer
Episode	Square Tomatoes		
Episode	Tomatoes	Created	1977-04-19
Contributing Organization	NewsHour Productions (Washington, District of Columbia)	Asset type	
AAPB ID	cpb-aacip/507-x31ng4hp5t	Topics	Energy Agriculture Food and Cooking
		Rights	Copyright NewsHour Productions, LLC. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International Public License (https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode)

Record

Monroe Duncan



[Transcript](#) Not Available

If you have more information about this item than what is given here, we want to know! Contact us, indicating the AAPB ID (cpb-aacip/387-70zpcgn9).

Segment	Monroe Duncan	Description	No description available
Contributing Organization	WHRO (Norfolk, Virginia)	Asset type	
AAPB ID	cpb-aacip/387-70zpcgn9	Genres	Unedited
		Topics	Food and Cooking
		Media type	Moving Image
		Duration	00:20:00?
		Citation	Chicago: "Monroe Duncan." WHRO. American Archive of Public Broadcasting

Record

New Jersey Nightly News; Tomato Farmer Burl. Co Tape 1 10-6-82 / Tomato Farm



Transcript Not Available

If you have more information about this item than what is given here, we want to know! Contact us, indicating the AAPB ID (cpb-aacip/259-9w090c5t).

Series	New Jersey Nightly News	Description	No Description
Raw Footage	Tomato Farmer Burl. Co Tape 1 10-6-82 / Tomato Farm	Broadcast	1982-10-06
Contributing Organization	New Jersey Network (Trenton, New Jersey)	Genres	News
AAPB ID	cpb-aacip/259-9w090c5t	Topics	News
		Media type	Moving Image
		Duration	00:20:00?
		Citation	Chicago: "New Jersey Nightly News; Tomato Farmer Burl. Co Tape 1 10-6-82

Record

The Emily Rooney Show, WGBH Radio



Transcript

Not Available

If you have more information about this item than what is given here, we want to know! [Contact us](#), indicating the AAPB ID (cpb-aacip/15-j96057dg9n).

Series The Emily Rooney Show

Title WGBH Radio

Contributing Organization WGBH (Boston, Massachusetts)

AAPB ID cpb-aacip/15-j96057dg9n

Description Emily Rooney Show

Asset type Program

Topics Public Affairs

Publisher WGBH Educational Foundation, Publisher

Media type Sound

Duration 01:00:00

Citation Chicago: "The Emily Rooney Show, WGBH Radio," WGBH, American Archive of Public Broadcasting (WGBH and the Library of Congress), Boston, MA and Washington, DC, accessed September 20, 2018.
http://americanarchive.org/catalog/cpb-aacip_15-j96057dg9n.

MLA: "The Emily Rooney Show, WGBH Radio." WGBH, American Archive of

Wide variety

- Of content from locations across the country
- Lots of different types of speakers
- Different speech patterns and accents
- 40 states
- politicians, academics, man on street, youth, elders, artists
- north, south, mid-west, west, east
- Creole, French, Caribbean Spanish, Alaskan native language
- Music and other non-spoken sound

- Computational Tools
- Speech-to-text
- Audio analysis
- Image Analysis
- Visualization of Data

How can we use them?

the potential:
transforming content into data



AUDIO ANALYSIS

High Performance Sound Technologies for Access and Scholarship (HiPSTAS)

generating additional programmatic data about the audiovisual material beyond the actual words being spoken in the recordings

- speaker identities
- applause
- laughter
- musical interludes

4,000 hours



The University of Texas at Austin
School of Information



The University of Texas at Austin
School of Information

speaker labeling

- WGBH-provided training set (all audio) 373.3 hours [speakers we aren't using etc.]
- Extended corpus (all audio) 951.8 hours [collected on own, keywords, etc. might change some of it not; some included in above #]
- "Haystack" test set (all audio) 494.2 hours [gave us to label because we think they're in there; might be some duplicated so we don't use it for training; use it to run the machine. kept on separate]
- Human-labeled speaker segments (all) 103.8 hours [everything]
- UBM_600 corpus 17.9 hours [all labeled speakers they gave us; plus PennSound UBM] Universal Background Model

HiPSTAS AAPB Output

AAPB speaker identity models
... so far

Hillary Clinton
Bill Clinton
James Baldwin
Malcolm X
Martin Luther King Jr.
Julia Child
Richard Nixon
Ronald Reagan
Lyndon Johnson
Gloria Steinem

Workshop
Process evaluation



The University of Texas at Austin
School of Information

Results

- We didn't get very much identified
- Really hard to do
- Took a lot of time to identify 1 speaker
- Translating the work from the research at UT Texas and tool training to our own catalog was undefined

SPEECH TO TEXT RESULTS

The Lexicon

- A list of words with their phonetic pronunciations
- The output of the speech recognizer will only include words that are part of the lexicon
- Words can have multiple pronunciations
- The lexicon links the acoustic model and the language model

archive AA R K AY V

hello HH AX L OW

hello HH EH L OW

pop P AA P

wgbh D AH B AX L Y UW JH IY B
IY EY CH

world W ER L D

up AH P



Search

help

referral program

my collections

casey davis

Pop Up Archive - NH-PR Presidential

Edit item

Delete item

AA-NHPR-presidential/cpb-aacip-503-zk55d8pb1z_NHPR95113.mp3

Suggested

Clear Suggested

- Keene ✓
- America ✓
- New Hampshire ✓
- Washington ✓
- United States ✓
- Mayor ✓
- teacher ✓
- Manchester ✓
- First inauguration of John Adams ✓
- United States Presidential inauguration ✓
- Bill Clinton ✓
- Paris ✓
- Newark State College ✓
- Governor ✓
- health insurance ✓
- First lady ✓
- Energy ✓
- Pat Russell ✓
- Henry Busham Clinton ✓
- Federal Bureau of Investigation ✓
- Holbrook ✓
- Newark State College ✓
- Released products ✓
- New President ✓
- United States Congress ✓

Item ID: AA-NHPR-presidential/cpb-aacip-503-zk55d8pb1z_NHPR95113.mp3

Duration: 00:38:57

Hide info

Premium Transcript

Search

Export

Save Text

Collapse

0:38:57 MP3



⏮ Shift+Tab Play from start of file

⏸ Tab Pause/Play

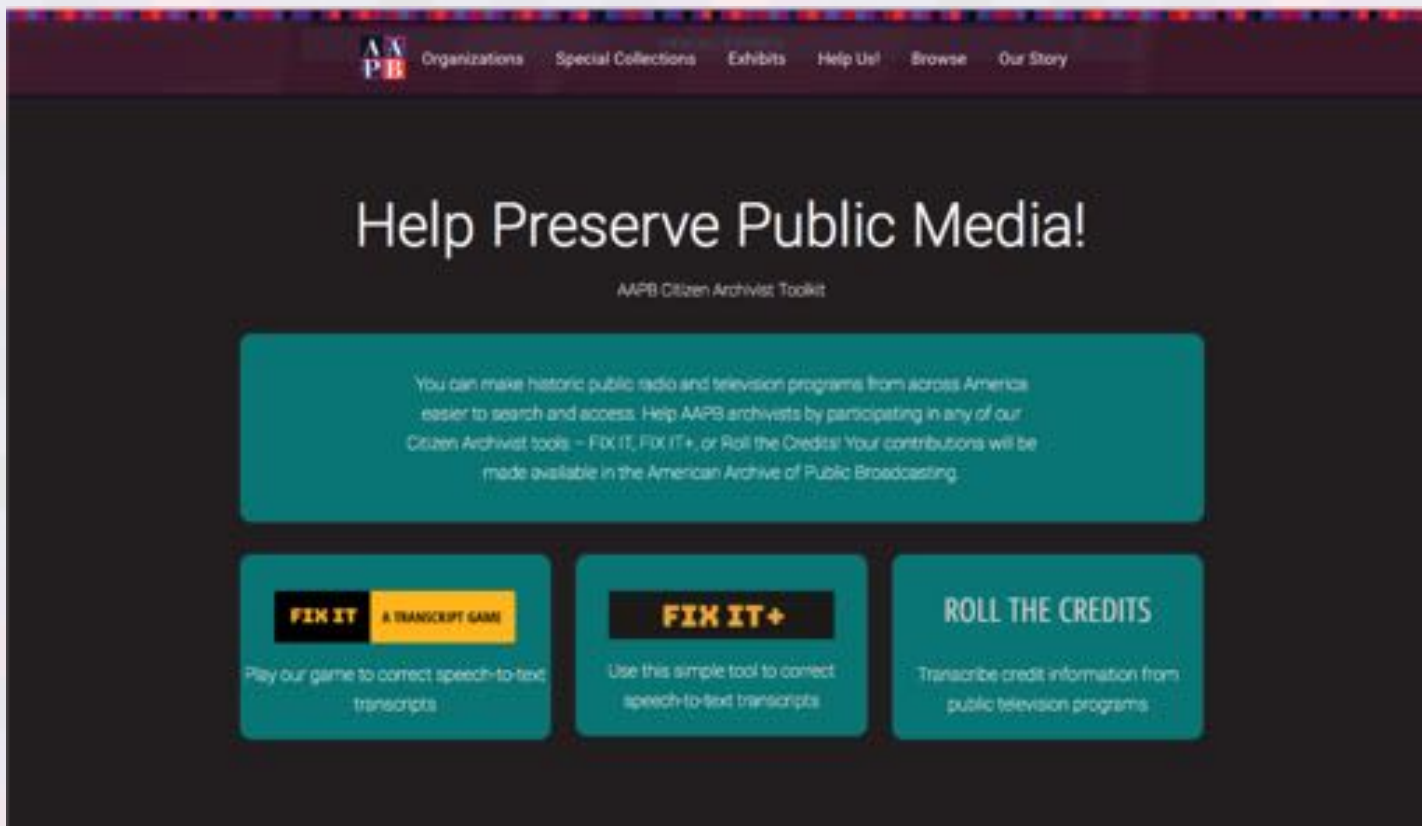
The Speech to Text Pipeline

- Task: given audio signal as input, choose the sequence of words that is the best match
- Two components to the model
 - Acoustic model (AM): which phonemes most closely match the signal
 - Language model (LM): which sequences of words are most likely
- We use Kaldi, an open source toolkit for automatic speech recognition (ASR)
- During the process of decoding the audio signal, Kaldi and recognizers store a lattice with hypotheses of the output for each utterance. Hypotheses with lower probabilities get "pruned" out
- At the end of the decoding process, the top hypothesis gets written to the output file

Transcript Creation Continues

- As part of the project, Pop Up Archive released updated language models for Kaldi.
- The code has been “Dockerized” by our University of Texas partners.
- We have incorporated transcript creation into our workflow as the collection grows.
- Github: <https://github.com/WGBH/kaldi-pop-up-archive>

Crowdsourcing tools



The screenshot shows the AAPB Citizen Archivist Toolkit website. At the top is a dark navigation bar with the AAPB logo and links for Organizations, Special Collections, Exhibits, Help Us!, Browse, and Our Story. The main heading is 'Help Preserve Public Media!' with the subtitle 'AAPB Citizen Archivist Toolkit'. A teal box contains a paragraph about making historic public radio and television programs easier to search and access by participating in Citizen Archivist tools. Below this are three teal boxes, each with a yellow header and a description of a tool: 'FIX IT A TRANSCRIPT GAME', 'FIX IT+', and 'ROLL THE CREDITS'.

AAPB Organizations Special Collections Exhibits Help Us! Browse Our Story

Help Preserve Public Media!

AAPB Citizen Archivist Toolkit

You can make historic public radio and television programs from across America easier to search and access. Help AAPB archivists by participating in any of our Citizen Archivist tools – **FIX IT**, **FIX IT+**, or **Roll the Credits**. Your contributions will be made available in the American Archive of Public Broadcasting.

FIX IT A TRANSCRIPT GAME

Play our game to correct speech-to-text transcripts

FIX IT+

Use this simple tool to correct speech-to-text transcripts

ROLL THE CREDITS

Transcribe credit information from public television programs



UNDER REVIEW

Zooniverse Project – “Roll the Credits”

ROLL THE CREDITS

ABOUT

CLASSIFY

TALK

COLLECT

FEEDBACK



What type of text appears in this frame?

- ☐ Only One Complete Credit Block
- ☐ Multiple Complete Credit Blocks
- ☐ Copyright Statement
- ☐ Other Text
- ☐ Multiple Types of Text
- ☐ No Text/Only Partial Credit Block(s)

You should sign in

Need some help with this task?

 Preferences

GAME MENU



KarenCariani

Total Points:

5130

1

IDENTIFY ERRORS

2963

Points

PLAY

2

SUGGEST FIXES

400

Points

PLAY

3

VALIDATE FIXES

1767

Points

PLAY

Share:

[Terms of Use](#)

FIX IT+ 4 Contributors 1984 James Baldwin, Jr. Home About Help Feedback

A Conversation With James Baldwin; James Baldwin Interview

This older version of "A Conversation With James Baldwin" replaces his earlier interview with Baldwin from "Perspectives, Negroes and the American Experience."

Duration: 0:09
[Listen to the audio version \(MP3\)](#)

Instructions

The following transcript was automatically generated using speech-to-text software, so there are probably errors. This tool will allow you to listen while you edit the transcript. For your convenience (see below), the audio will automatically pause after each line.

Use the redaction shortcuts or buttons in the toolbar below to reorganize the transcript and audio. As your edits will be automatically saved, but you will be immediately notified by others. Once enough people agree on the text of a particular line, that line will be visible to all and no longer editable.

The transcript contains lines that are "completed" ~~(highlighted in green)~~ which means they have been corrected by others and can no longer be edited. But you can still listen to them.

[View A Tutorial](#) [Get Started](#)

0:00 One of the significant things about the present revolution of the Negro
 0:04 people in America, is maybe the fact that for the first
 0:08 time
 0:12 there is genuine communication between negroes and whites.
 0:16 Negroes are saying out loud now
 0:20 things which they have long said only to themselves.
 0:24 (Rustling sound)
 0:28 Probably one of the most articulate,
 0:32 passionate, and clear communicators
 0:36 to the American conscience,
 0:40 is my guest, James Baldwin.
 0:44 James Baldwin's name is known throughout America
 0:48 for saying so passionately, and so

[Play](#) [Stop](#) [Next Line](#) [Previous Line](#) [Play Next](#) [Next Line](#) [Previous Line](#) [Next Line](#) [Previous Line](#)

FIX IT+

FIXITPLUS.american
archive.org

0:09 ☒ Can you tell me more?

☐ Yea- yeah care to start

0:15 ☐ why do want to show

An example of editing a transcript

NYPL Community Oral History

transcribe.oralhistory.nypl.org/transcripts/adam-payne-idujw

Casey E. Davis

The New York Public Library
Community Oral History Project

transcriptEditor

Adam Payne

Browse

Search

About

FAQ

Tutorial

Track My Progress

Show Conventions

00

✓

👤

Hello my name is Monica Diaz and I'm going to have a

05

✓

👤

conversation with Adam Payne. ADAM PAYNE: Hi. MONICA DIAZ: Hello, for visible

10

✓

👤

an Oral History Project at the New York Public Library it is

15

✓

👤

is February eighteenth two thousand fifteen and this is being recorded at the

20

✓

👤

the Inwood Public Library in New York. ADAM: It is. MONICA: Thank you so

25

✓

👤

so much

25

✓

👤

Adam. You know can we, can we start off by having you

31

✓

👤

Tell me a little bit about yourself.

33

✓

👤

Well, my name's Adam Payne, um

37

✓

👤

I was born a long, long time ago in a galaxy far, far away.

43

✓

👤

actually I was born in 1971 in Cambridge, Mass.

Play

⌘ + ⏮

⏮

Next Line

1

OR

⏮

Prev Line

1

⏮

Prev Word

⌘ + ⏮

⏮

Next Word

⌘ + ⏭

⏭



```
{
  "language": "en-US",
  "Siri": {
    "parts": [
      {
        "id": 24387886,
        "text": "A production of the Mississippi Center for Educational Television. Number three series a conversation",
        "start_time": "6.13",
        "end_time": "10.97",
        "speaker_id": 889358
      },
      {
        "id": 24387887,
        "text": "with program 28 30 day to direct.",
        "start_time": "10.97",
        "end_time": "37.78",
        "speaker_id": 889358
      },
      {
        "id": 24387888,
        "text": "You you.",
        "start_time": "37.78",
        "end_time": "53.35",
        "speaker_id": 889359
      },
      {
        "id": 24387889,
        "text": "MARTIN Carthy metal smith and Potter for",
        "start_time": "53.35",
        "end_time": "57.94",
        "speaker_id": 889360
      },
      {
        "id": 24387890,
        "text": "Marigold Mississippi they are a",
        "start_time": "57.94",
        "end_time": "61.89",
        "speaker_id": 889360
      },
      {
        "id": 24387891,
        "text": "native of the Internet",
        "start_time": "63.45",
        "end_time": "68.42",
        "speaker_id": 889360
      },
      {
        "id": 24387892,
        "text": "you know. I know",
        "start_time": "68.7",
        "end_time": "73.2",
        "speaker_id": 889360
      },
      {
        "id": 24387893,
        "text": "we know and what the noted European are",
        "start_time": "73.68",
        "end_time": "78.55",
        "speaker_id": 889360
      },
      {
        "id": 24387894,
```

- 0:06 [] A conversation in the Mississippi Center for Telemedicine. Thera Ham... A conversation...
- 0:11 [] conversation with Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 0:38 [] A conversation with Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 0:53 [] what Thera Ham... is the McCarly... Thera Ham... Thera Ham...
- 0:58 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- Line has been completed and is no longer editable
- 1:03 [] native Mississippi, born in Paris, Mississippi. Internationally known.
- 1:09 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 1:14 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 1:19 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 1:23 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 1:30 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 1:35 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 1:40 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 1:44 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 1:48 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 1:53 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 1:58 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 2:03 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 2:08 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 2:13 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 2:18 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 2:23 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 2:28 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 2:33 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 2:38 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 2:43 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...
- 2:48 [] Thera Ham... Thera Ham... Thera Ham... Thera Ham...

```
language: ""
id: 178
parts:
  0:
    id: 31602
    text: "A production of the Mississippi Center for Educational Television. Take number three. A series conversation"
    start_time: 6.129
    end_time: 10.97
    speaker_id: 1
  1:
    id: 31603
    text: "program with Theora Hamblett. Length 28 minutes 30 seconds. Date 12/16/75. Director Seymour."
    start_time: 10.97
    end_time: 37.78
    speaker_id: 1
  2:
    id: 31604
    text: "A conversation with Theora Hamblett. Talking to today with Mississippi "
    start_time: 37.78
    end_time: 53.35
    speaker_id: 1
  3:
    id: 31605
    text: "artist Theora Hamblett is Lee McCarty. Metalsmith and Potter from"
    start_time: 53.35
    end_time: 57.939
    speaker_id: 1
  4:
    id: 31606
    text: "Marigold Mississippi. \"the Art of Theora Hamblett\""
    start_time: 57.939
    end_time: 61.89
    speaker_id: 1
  5:
    id: 31607
    text: "native Mississippian, born in Paris, Mississippi. Internationally known"
    start_time: 63.45
    end_time: 68.42
    speaker_id: 1
  6:
    id: 31608
    text: "Painter. You might be interested to know"
    start_time: 68.7
    end_time: 73.2
    speaker_id: 1
  7:
```

Record

En Camino, 205; Special: Juan Serrano, Master Guitarist



Transcript Available

If you have more information about this item than what is given here, we want to know! Contact us, indicating the AAPB ID (cpb-aacip/88-655dv9t7).

Series En Camino**Episode Number** 205**Episode** Special: Juan Serrano, Master Guitarist**Contributing Organization** KRCB (Rohnert Park, California)**AAPB ID** cpb-aacip/88-655dv9t7**Description**

A special interview and on stage performance by Juan Serrano, Master Guitarist, who tells about his origins, life and introduction to music and the Art of Flamenco. He began his professional life at thirteen years of age. The art of Flamenco is described as a mix of dance, guitar music, counter rhythms of hand clapping, and 'el duende' which are the deep emotions expressed by the artists.

Created

1987-06-11

Asset type

Episode

Topics

Music
Performing Arts
Education
Spanish Language

Creator

KRCB FM, Producing Organization
The San Francisco Foundation, Co-Producer
Gaona, Francisco, Co-Producer
Lukens, Jane, Co-Producer

Contributor

Serrano, Juan, Interviewee
Serrano, Juan, Performer

Kaldi Output for Music



Lord thank you. Lord. You're through. You're.
Through. Mm mm. Mm. Mm mm mm mm mm.
Mm. Mm. Laura Iraq of. Glory. From. From.
From. Three three. Three. Three. Three. Three.
Three.

Laura it thank. You. That was it. And those are
the good other than those of the other. But
other than that I'm.

But you. Know. In PSA jogging for the leg. That.
Said alone but not for what they love that about
them. In the U.S. I don't really like deep down
this guy John. Has seen me naked young. You're
that one or put up or go to a lie even. When
we're going to be in a way that it was. A scene or
so because John you and I will note that up why
go to a lie. When we're going to be in before and
then they run a scene will see My God. You up
or something. On Monday another tour and yet.
People are coming up a lot of. Joy to make us
even more. Of this season might. Be made
upon home with us so we're left with.

The money to go before the scene was bought.

Kaldi Output for Helicopter B-Roll



A.

Leg.

Length.

Length.

Eh I'm

good.

Thank you.

Kaldi Output for Spanish Language



You mean you think you know always going. In approximately I see that I'm being in a city that the community is Brucey putting I mean in and deal with it and they see it. Are they looking for. And COMINO is improved on so many ideas to explore that they must be in for fancier graphic thing like that Latina. Woman That is keep it down for a. Made up out of there. I mean the older you know the moment they do it and I mean no no shangri la I wound up pretty compressed. But then finally the. Most famous presenter being asked about Odama seeking

impact done like a 20 year old uncle. I miss my foresight when local news reported a loss for him which was the original being with you we. Let's see it. No problem I say to you since it was your own this and not a Californian. When you guys buy up our land Mr. Merhi Nera for law enforcement you said to be sure that will cost them. Less personally and COMINO is bloviated not very effective. They must be telescreen are so fake now. But up Roamer they don't know about this the past few months I feel I am sure that. If we didn't know what it is as soon as one has been able to from my strong ally in cancer they mine. I say most I would have had lots of them in Dallas. I remember with Ahmadinejad. See that he's a see the. This was very programa I said Think end up in the Mystica. I think can assume that I would be in the family is in Gaza you know of. Just a movement that eluting mother in that race in Illinois. I mean these people are Latino and they they face. Like the Latinos in going through all of Emmelina repeat them because they see areas because it reduces Saulteaux elated.

Kaldi Output for French-Creole Language



All.Right. OK. OK. West Virginia Virginia.
Plus. Brown. Girlfriend try a temporary
quick. Corollary proper and I pointed to a
song. Made on your apartment made I'm
easy. I generally. Don't jump to a long.
Record. Your point here already called an
average storm either calling. Me a fearful
planner for us are really powerful for one I
don't live for or see me privately just 75
miles from or see. Them. President
Reagan National Award. You have more
screen. Generation. We have a long trip.
For. A trip. Do you see music music from
music. And there is a common sense for
him when the. Nazi party. No no can one.
Man may or. Yet you monkey have made
of the tears and. Some are. Good only the
Starkey blues the silly your so I look at this
hour. Mr Parker knew me or started. Their.
Own eyes. What I am. Not very good.


We know nothing about this item.

Record

Waves Of Wisdom

Transcript

Not Available



A video player interface showing a portrait of an elderly man. He is wearing a dark flat cap, dark sunglasses, and a blue and white zip-up jacket. The video player has a progress bar at the bottom with a play button, a volume icon, and a time display of -19:23.

If you have more information about this item than what is given here, we want to know! Contact us, indicating the AAPB ID (cpb-aacip/127-03cz8zdq).

Title	Waves Of Wisdom	Description	No description available
Contributing Organization	KYLK (Bethel, Alaska)	Genres	Unedited Interview
AAPB ID	cpb-aacip/127-03cz8zdq	Media type	Moving image

Kaldi Output for Yup'ik Language



Your all right already. Yeah OK. For the record.
Speaking to James guy Greek looked him up in the crowd doesn't mean we thought i hope I do I'm trying to talk but I think we shall I do you can and we shall not be in hall and you know

I do wish I could. I will call that good old myth or the old call wouldn't let me out with a dog. Be sure they left the water good. How does how do you lead that cuckoo clock in the south have you genius. And I don't miss out on your good

guy you're not familiar with what do I do when you need to whom you are now there. Dorise you I laughed and I dunno it still gets you can pick on a new record. Yeah gotta be deader than a good guy. Let's go on a con not go off on your own of kind and how can you know how God left her daughter. Julie I don't know we don't really care what you never knew that people can't go through the muck of the kind in which our not. I don't come out now if you're going to cause you need no with no trade in Kenya and it could be should be a good deal or not a number could not go home now or that not gotten bad bitches want to go Honey you do not want me one culture not clear what we've got to do what you want to try New Delhi new kind o woman in her thought didn't think enough on line to get rid of they have a machine like you.

Kaldi Output for Spoken English

- Approximately 81% word accuracy rate
 - not including punctuation errors
- Examples:
 - 95% accurate for 1960s radio program from Boston (no accents, one speaker)
 - 55% accurate for 1970s television program from Mississippi (strong Southern U.S. accent)

Item 1
4,497 words, 91% matched

The following is a production of the public affairs department of	« 12 words »	The following is a production of the public affairs department of
W will be be. Louisiana		WLPB. Louisiana:
the state we're in. With public affairs director Beth George. Good evening. Thank you for joining us on this first edition of the Louisiana the state	« 26 words »	The State We're In. with Public Affairs director, Beth George. Good evening. Thank you for joining us on this first edition of the Louisiana: The State
weary and		We're In.
tonight we examine with our guests and on film the state prison at Angola. We have with us this evening three knowledgeable guests. We have Mr. C. Paul Phelps	« 29 words »	Tonight we examine with our guests and on film the state prison at Angola. We have with us This evening three knowledgeable guests. We have Mr. C. Paul Phelps,
was		the
secretary and head of the Department of Corrections. We also have with us Dr. Mark	« 15 words »	secretary and head of the Department of Corrections. We also have with us Dr. Mark
Carlton		Carleton,
research director for the Public Affairs Research Council former professor of history at LSU and author of a book on the history of Louisiana's penal system. And finally we have with us this evening Representative Bobby Freeman of	« 38 words »	research director for the Public Affairs Research Council, former professor of history at LSU, and author of a book on the history of Louisiana's penal system. And finally, we have with us this evening Representative Bobby Freeman of
plaque and		Plaquemine,
a Louisiana state representative who is one of Governor Edwards' chief	« 11 words »	a Louisiana State Representative, who is one of Governor Edwards' chief
		floor
leaders in the house and a member of the House Appropriations Committee. Angola has been in the news almost constantly since June 1975 when federal district judge	« 27 words »	leaders in the House and a member of the House Appropriations Committee. Angola has been in the news almost constantly since June 1975 when federal district judge
		E.
Gordon West issued a court order declaring the prison overcrowded and the living	« 18 words »	Gordon West issued a court order declaring the prison overcrowded and the living

Types of Errors Corrected by Users

- Station call letters
- Mis-transcription of words spoken in southern accents, e.g., “weary and” vs “we’re in”
- Local town names, e.g., “plaque and” vs. “Plaquemine”
- Person names, e.g., “Laurence” vs. “Lawrence”
- Numbers spelled out vs. numeric
- Adding words completely missing from original transcript
- Incorrect “corrections” by crowdsourcing participants, e.g. “achieved” vs. “achievedd” in the “corrected” transcript

JSON transcripts are stored on AAPB's Amazon S3 account

Transcripts are indexed for keyword searching on the AAPB website

Transcripts are made available alongside the media on the record page

Transcripts can play as captions within the player

Transcripts can be harvested via an API and used as a dataset for research such as a digital humanities project

once corrected...

Record

A Conversation With; Theora Hamblett



Transcript

Search...



Hide

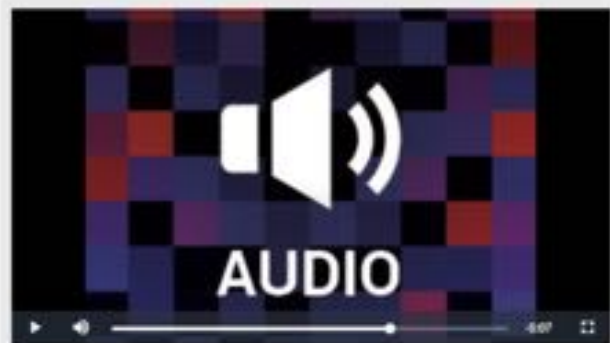
- A production of the Mississippi Center for Educational Television. Take number three. A series conversation program with Theora Hamblett. Length 28 minutes 30 seconds. Date 12/16/75. Director Seymour. A conversation with Theora Hamblett. Talking to today with Mississippi artist Theora Hamblett is Lee McCarty, MetalSmith and Potter from Marigold Mississippi. "the Art of Theora Hamblett" native Mississippian, born in Paris, Mississippi. Internationally known Painter. You might be interested to know though we know in this country who Theora Hamblett is. What the noted European art critic and historian, Otto Behamarian had to say in his book Modern Primitives - Master of Naive Paintings
- And there is Theora Hamblett with one of her glaring yellow visions quite sufficient of itself to serve as flashbulbs for the photographers who would surely not miss this opening. The opening they were referring to was the world exhibition in Brussels. The pane they were referring to was

If you have more information about this item than what is given here, we want to know! Contact us, indicating the AAPB ID (cpb-aacip/60-81wdc1hh).

Series	A Conversation With	Description	111. QBTf. Theora Hamblett (1895_) African-American primitive painter from Mississippi.
Episode	Theora Hamblett		A Conversation With is a talk show featuring discussions with public figures in Mississippi.
Contributing Organization	Mississippi Public Broadcasting (Jackson, Mississippi)	Asset type	Episode
AAPB ID	cpb-aacip/60-81wdc1hh	Genres	Talk Show
		Topics	Fine Arts
		Media type	Moving Image
		Duration	0:28:30

Record

WGBH Journal; Adrienne Rich



Transcript

weather



Hide

he would not exclude from participation in the discussion



any Palestinian element or any ultimate choice of governance. Well we may hear more of this or some variant of it. Now that it's over the week carried New England **weather** beyond the conversational stage. A phenomenal week starting with the heaviest snow since **weather** records began, then torrential rain to turn snow banks into flooding. And then yesterday, a May day of temperatures that would have brought the tulips up if they could have found their way through the soggy snow. And to top it off, freezing again, as our **weather** today swings back to a January norm. Now we shiver at the news of blizzards that have paralyzed the whole middle of the country. Floundering through such a week, has blinded our interest to most other news. And preoccupation with the **weather** may have made it easier for



the government to hold back information about the disintegration of a

If you have more information about this item than what is given here, we want to know! Contact us, indicating the AAPB ID (cpb-aacip/15-05s7hb9).

Series WGBH Journal

Episode Adrienne Rich

Contributing Organization WGBH (Boston, Massachusetts)

AAPB ID cpb-aacip/15-05s7hb9j

Description

Poet Adrienne Rich, Louis Lyons. Engineer: Carter

WGBH Journal is a magazine featuring segments on local news and current events.

Asset type

Segment

Genres

News
Magazine

CL experts

- Improving named entity vocabularies
- Forced alignment
- Time stamp for bars and tone
- Music identification
- Foreign language identification and transcription
- OCR of text on screen (lower thirds, credits)

CL Experts and Archivists

- A larger need for more accurate output and ease of use of computational tools for audiovisual archives to create descriptive metadata and annotations.

Help Us!

Access our Dataset

- Metadata API (OAI-PMH and PBCore)
 - <https://github.com/WGBH/AAPB2#api>
- Transcripts API – contact us to get credentials
- Media as a Dataset (MaaD) – contact us to get digital audio/video copies for computational research



<http://fixit.americanarchive.org>

#FixItAAPB



americanarchive.org

Karen Cariani

Karen_Cariani@wgbh.org



[@amarchivepub](https://twitter.com/amarchivepub)

Casey Davis-Kaufman

casey_davis-kaufman@wgbh.org



facebook.com/amarchivepub