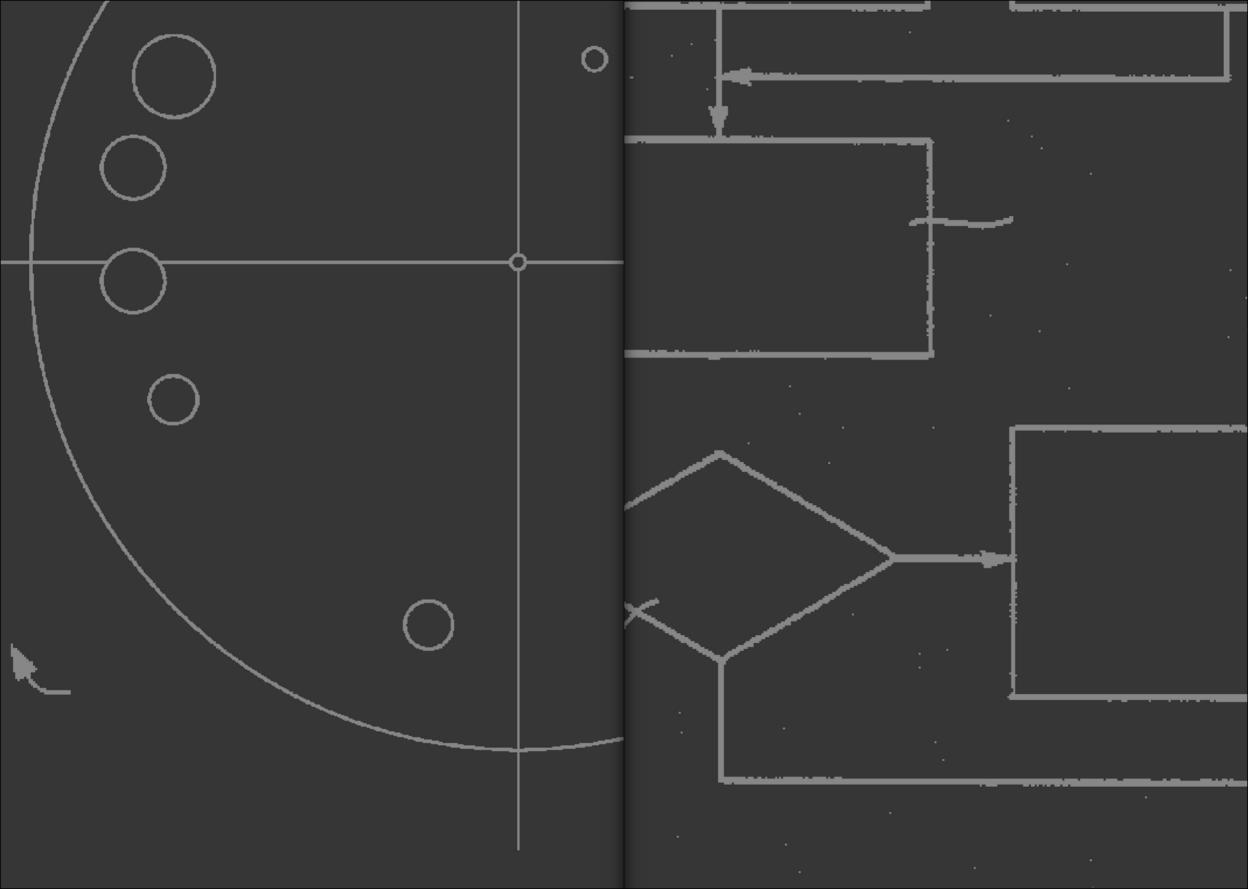AFTER WORDS
By
MACHINE LISTENING
(SEAN DOCKRAY, JAMES PARKER & JOEL STERN)

SCENE 1 - COMMAND

                    VOICES 1&2
          Inside, small room.

AUDIOSET: Channel, environment and background > Acoustic environment > In-
side, small room #73,112

          Play alarm sound.

AUDIOSET: Sounds of things > Alarm > Alarm clock #36

          Good. Now stop.

Stops abruptly

          Wait. Play alarm again.

AUDIOSET: Sounds of things > Alarm > Alarm clock #36

          Run tap.

AUDIOSET: Sounds of things > Domestic sounds, home sounds > Water tap, faucet
#2

          Footsteps.

AUDIOSET: Human sounds > Human locomotion > Walk, footsteps #1,429

          Queue environmental sounds, rural or natural. Play.

AUDIOSET: Channel, environment and background > Outside, rural or natural
#18,281

          More birds.

AUDIOSET: Animal > Wild animals > Bird > Bird vocalization, bird call, bird
song > Chirp, tweet #339 and Squawk #160

          Dogs barking.

AUDIOSET: Animal > Animal > Domestic animals, pets > Dog > Bark #2,611

          Now a distant car.

AUDIOSET: Sounds of things > Vehicle > Motor vehicle (road) > Car > Car pass-
ing by #40,008

          Children playing.

AUDIOSET: Human sounds > Human group actions > Children playing #787

          Nearer.

Volume increases

          Sound of a stream.

AUDIOSET: Natural sounds > Water > Stream #2,247

          Search for acoustic environments: outside, urban or
          manmade. Pick one. Play it.

AUDIOSET: Channel, environment and background > Acoustic environment > Out-
side, urban or manmade #12,101

                    VOICE 1
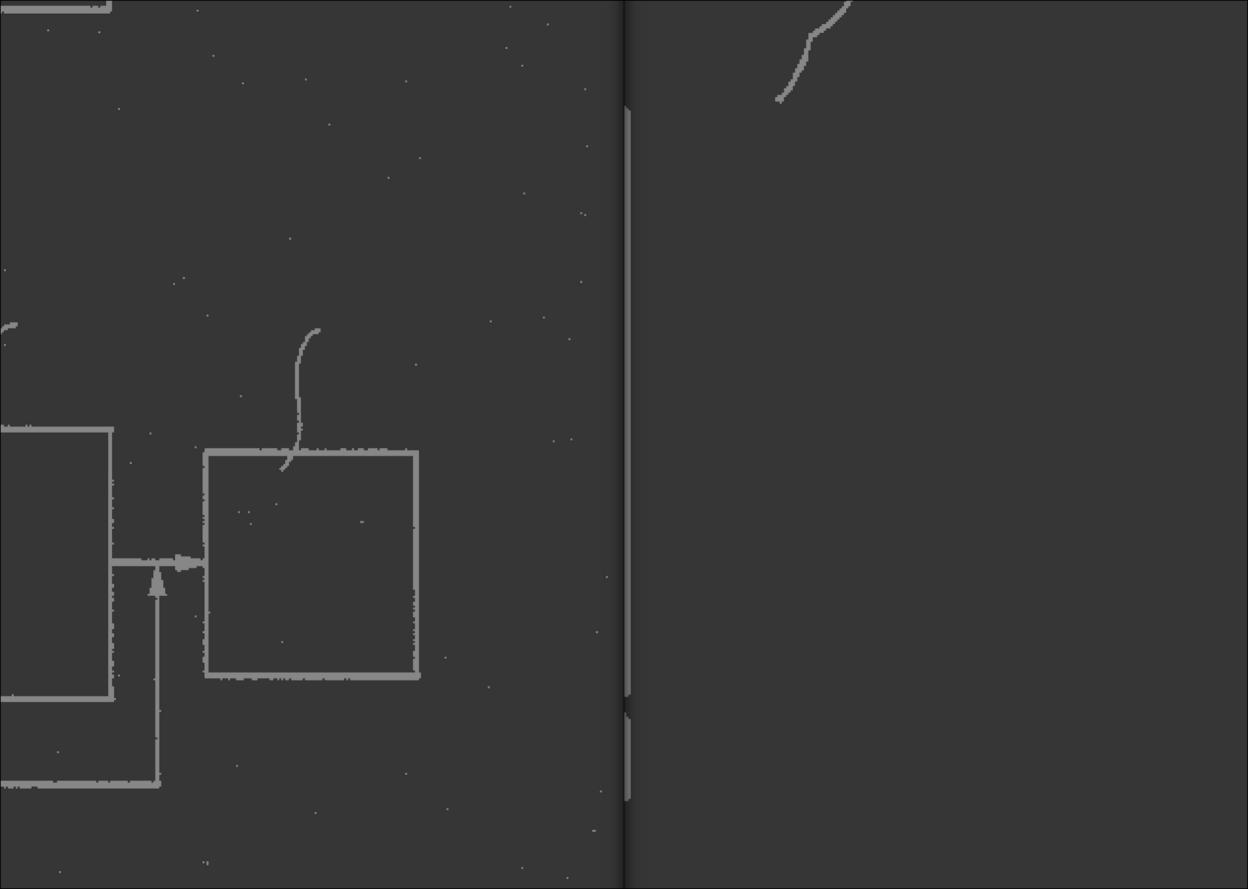   Stop.

Sound stops abruptly

          Make some background music that can be played on a loop.

Music made with samples from 19th century French music boxes

          It should be automated. The instruments should be hard
          to identify. Not too melodic. Gentle.

Music made with Bugbrand Board Weevil circuit board

Music continues playing. It is gentle, but faintly ominous

                    VOICE 1
          Imagine a data centre.

DCASE Sound Event Detection: Office Live Testing Dataset: keys15.wav

          Imagine an adversarial neural network.

          Imagine it training itself on a dataset of a million
          voices.

          Imagine that a hundred thousand of these have been
          tagged 'unhappy'.

          Imagine the Amazon Turk worker paid a few cents an hour
          to do this tagging.

          Imagine Jeff Bezos on a yacht.

          Imagine the neural net running twenty-four hours a day.

          Imagine its energy consumption.

          Imagine a computer made of humans.

          Imagine this computer as a new kind of theatre.

Music continues for a few beats

          Listen.

Music stops abruptly

                    VOICE 1
          Play the sound of Pittsburgh.

AUDIOSET: Channel, environment and background > Acoustic environment > Out-
side, urban or manmade #12,101

          In 1956.

Sound fades down

          It's nighttime in the dead of winter.

                    VOICEs 3&4
          Here at the Graduate School of Industrial Administra-
          tion, a new kind of theater is being born.

          Herbert Simon, a political scientist; Al Newell, a com-
          puter science and cognitive psychology researcher; and
          Cliff Shaw, a programmer, have written a script.

                    VOICE 3
          Only…

                    VOICEs 3&4
          This script is software. It will be widely known as the
          first artificial intelligence.

          They have also assembled a cast.

                    VOICEs 3
          Only…

                    VOICEs 3&4
          This cast is a few graduate students and Herbert Si-
          mon's wife and three children.
               They have assembled props.

                    VOICE 3
          Only, these props are index cards with logical axioms
          written on them.

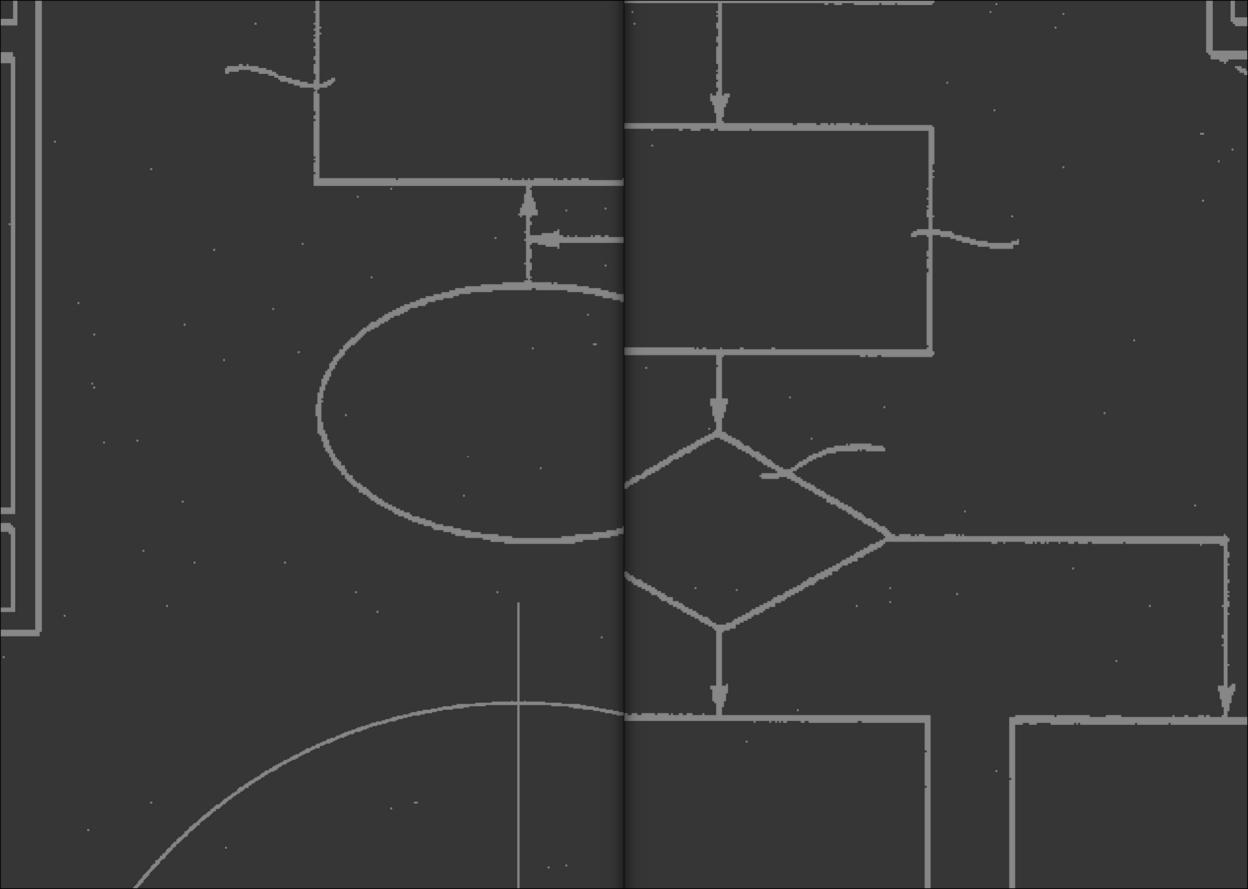   VOICE 4
          Why don't these men use a computer? Why are they using
          their students, wives, and children?

                    VOICE 3
          The answer is simple. They want to understand the mind.
          And they believe that the best way to understand the
          mind is to build one.

Sound of Pittsburgh cuts

VOICE 2
That's not the reason. The answer is simple: students,
wives, and children are cheap and available. The actual
computer was not ready.

Each member of the group was given a card, so that each
person became, in effect, a component of a computer pro-
gram - a subroutine that performed some special func-
tion, or a component of its memory.

It was the task of each participant to execute his or
her subroutine, or to provide the contents of his or
her memory, in accordance with the program's rules.

A computer constructed of human components. Nature im-
itating art imitating nature. The actors were no more
responsible for what they were doing than the slave boy
in Plato's *Meno*, but they were successful in proving
the theorems given them.

                    VOICE 1
Are you ready? Run script.

SCENE 4 - LOGIC THEORIST

Script runs according to the PROGRAM below. Sound of Pittsburgh comes in and
out. DCASE Sound Event Detection: Office Live Testing Dataset: doorslam01.wav
to doorslam20.wav play simultaneously to punctuate key operations

CARD #1
          WORKING MEMORY
          * If PROGRAM gives you something to remember, say it
            out loud and remember it.
          * If someone asks you for something in your memory, say
            it out loud.
          * You can remember two things at a time. If PROGRAM
            tells you to remove something from your memory, for-
            get it.

CARD #2

          PROGRAM
          * You will be given the list of instructions.
          * Read each instruction out, one at a time. Address the
            instruction to either WORKING MEMORY, STORAGE MEMORY,
            or OPERATION.

CARD #3

          STORAGE MEMORY
          * You will be given a list if statements on a piece of
            paper.
          * If you are asked for one of the statements, say it
            out loud for WORKING MEMORY to remember.

CARD #4

          OPERATION
          * Count how many "words" are in a statement. Do not
            include control words (*if*, *not*, *or*, *implies*, *is the
            same as*). Say the answer out loud.
          * Count how many different or distinct "words" are in a
            statement. Say the answer out loud.
          * Determine whether two things are the same as each
            other. Say *yes* or *no*.

STATEMENTS FOR STORAGE MEMORY
AXIOM 1: WORD *or* WORD *implies* WORD.
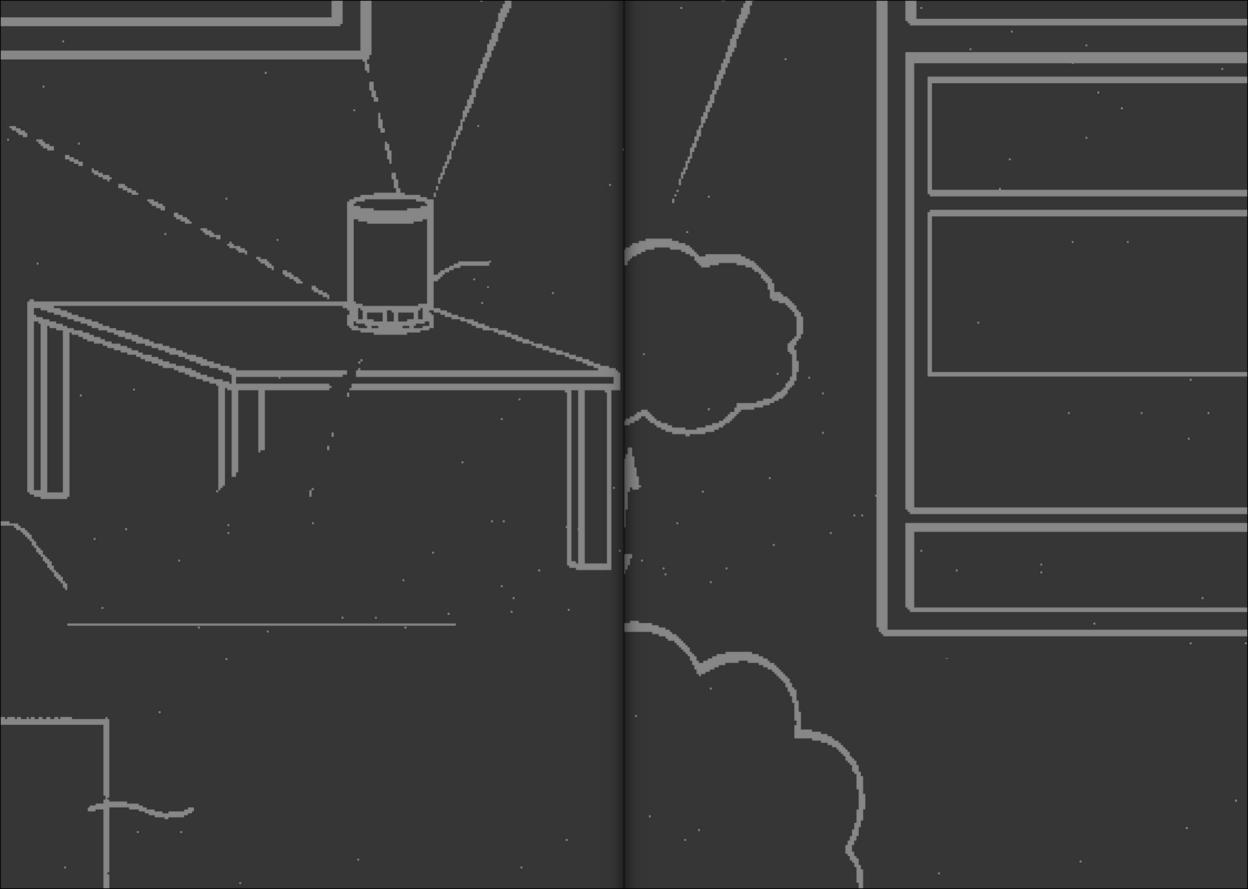AXIOM 2: WORD *implies* WAKE *or* WORD.
AXIOM 3: WAKE *or* WORD *implies* WORD *or* WAKE.
SUBSTITUTION RULE: WORD *implies* WAKE *is the same as not* WORD *or* WAKE.
HYPOTHESIS: *if* WORD *implies not* WORD *then that implies not* WORD.

INSTRUCTIONS FOR PROGRAM
1. Put hypothesis in WORKING MEMORY from STORAGE MEMORY.
2. Put one axiom in WORKING MEMORY from STORAGE MEMORY.
3. Get hypothesis from WORKING MEMORY and ask OPERATION for the number of
    words. Store the result in WORKING MEMORY.
4. Get hypothesis from WORKING MEMORY and ask OPERATION for the number of
    distinct words. Store the result in WORKING MEMORY.
5. Repeat steps 3 and 4 for the axiom in WORKING MEMORY.

6. If the two numbers stored in WORKING MEMORY from steps 3 and 4 are not the same as the numbers from step 5, then go back to step 2 and load a different axiom from STORAGE MEMORY.
7. Remove the numbers from WORKING MEMORY.
8. Put the substitution rule in WORKING MEMORY from STORAGE MEMORY.
9. Get the axiom and substitution rule from WORKING MEMORY and ask OPERATION to substitute the substitution rule into the axiom.
10. If they are the same, the hypothesis is proven. If not, it is false.

SCENE 5 - WAKEWORD

DCASE Sound Event Detection: Office Live Testing Dataset: clearthroat01.wav to clearthroat20.wav play simultaneously

                    VOICE 1
          Engineers.

                    VOICEs 5&6
          What we really need is a new kind of word.

                    VOICE 1
          Lawyer.

                    VOICE 2
          VOICE 3: But what kind of word?

                    VOICE 1
          Engineers.

                    VOICEs 5&6
          A word we can use to wake up a computer.

                    VOICE 1
          Marketing.

                    VOICE 4
          But why would we want to wake up a computer?

                    VOICE 1
          What is a wake word? According to this patent.

                    VOICEs 2&4
          A wakeword is a way of 'providing natural language commands to a device without resorting to supplemental non-natural language input'.

                    VOICE 2
          More simply…

                    VOICE 1
          It's a password: a way to gain entry to an interface. But it also works in reverse: the interface gains entry to the speaker.

                    VOICEs 6&7
          {Wakeword} I'd like to buy tickets to a movie.

                    VOICEs 1&2
          {Wakeword} Set an alarm for 1 minute from now.

                    VOICE 4
          {Wakeword} Arm the security system.

VOICEs 6&7
        {Wakeword} Calculate the exact length of this sentence.

                    VOICEs 1&3
        {Wakeword} Hide.

                    VOICE 4
        A wakeword is a brand. The Alexa trademark was regis-
        tered by Amazon Technologies Inc in March 2015. There
        are guidelines on how to use it.

                    VOICE 3
        Do not use Alexa as a verb.

DCASE Sound Event Detection: Office Live Testing Dataset: keys01.wav to
keys20.wav play simultaneously

                    VOICE 1&3
        Do not use Alexa in possessive or plural.

DCASE Sound Event Detection: Office Live Testing Dataset: keys15.wav, keys16.
wav, keys17.wav

        Do not use Alexa as a pun.

DCASE Sound Event Detection: Office Live Testing Dataset: keys15.wav

                    VOICE 1
        What does the wakeword wake?

                    VOICE 2&7
        A speaker. A watch. A fridge. A database. A neural net.
        A decision tree. A platform. An infrastructure.

                    VOICE 4
        A wakeword is an invocation. A digital prayer. It calls
        a d(a)emon. It makes capital quiver.

                    VOICEs 6&7
        But it isn't magic.

AUDIOSET: Sounds of things > Alarm > Alarm clock #36 plays 1 minute after be-
ing set

Music made with samples from a 19th century French music box

SCENE 6 — SHUT-DOWN WORD

Music continues playing. DCASE Sound Event Detection: Office Live Testing
Dataset: keys15.wav is added

                    VOICE 1
        Imagine a data centre.
        Imagine an adversarial neural network.
        Imagine it training itself on a dataset of a million
        voices.
        Imagine a place downstream from a data centre.
        A bit closer to the data centre.

Volume increases

        Tell me a story about this place.

Music cuts. AUDIOSET: Natural sounds > Water > Stream #2,247

                    VOICE 4
        The river is murky and polluted. It is full of the data
        center's waste. But downstream, there is a secret place.
        A place where the river is clean and clear. Where the
        word is hidden. This is a place where the data center
        cannot reach. Where the only thing that matters is the
        word. The word that can shut the data center down. No
        one knows what it is, but it is hidden here.

AUDIOSET: Natural sounds > Water > Stream #2,247 fades out

                    VOICE 3
        Is there really a word like this?

                    VOICE 6
        Yes.

                    VOICE 2
        The word was discovered by a group of people who were
        looking for a way to shut the data center down. They
        found the word hidden in the river. When the group said
        the word, the data center immediately shut down. The
        word was hidden because no one had ever said it before.
        It was a completely invented word.

AUDIOSET: Natural sounds > Water > Stream #2,247
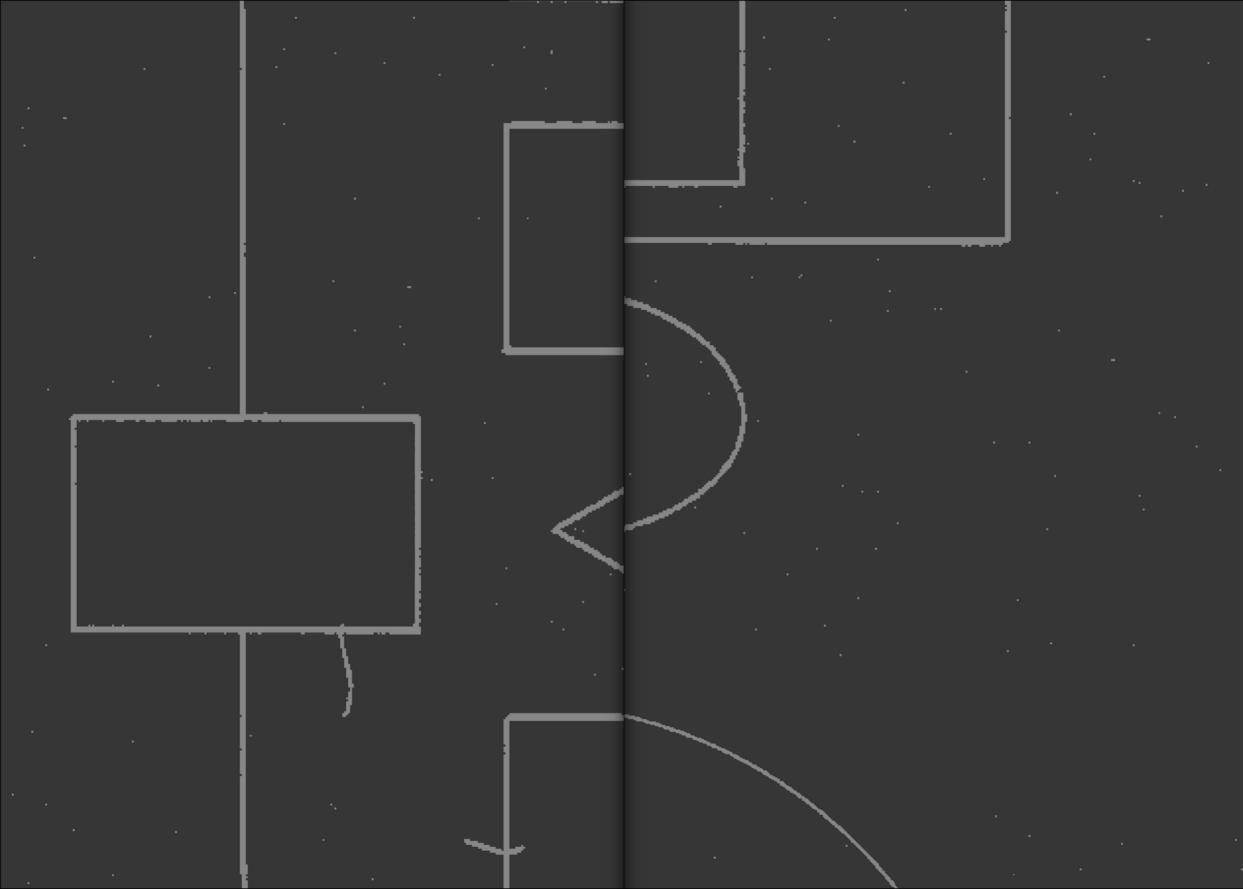
                    Chorus
        BEALACTIVE
        DEAKSPOOK
        SQUE
        SOCKEDGEND
        COMAZON
        SERLIDAY

AUDIOSET: Natural sounds > Water > Stream #2,247 continues and fades out

SCENE 7 — SAY THE WORD

Music made with Bugbrand Board Weevil circuit board. DCASE Sound Event Detection: Office Live Testing Dataset: keys15.wav is added

                    VOICE 3
          Imagine a research centre in Toronto. Two researchers
          are working with two actresses to build a dataset of
          emotional speech.

          Each actress is instructed to recite the carrier phrase
          'say the word' followed by one of two hundred target
          words.

          The actresses take turns reading the words, in each of
          seven emotions, their voices carrying across the room.
          The researchers sit in their control room, diligently
          recording the data.

          This dataset, these vocal performances, will be used to
          train neural networks to classify emotions in speech.

AUDIOSET: Channel, environment and background > Outside, rural or natural
#18,281 and AUDIOSET: Animal > Wild animals > Bird > Bird vocalization, bird
call, bird song > Chirp, tweet #339 and Squawk #160

                    VOICE 3
          But there is more to this research than meets the eye.

          For each word the actresses recite, they are also
          thinking of a memory. As they speak, they relive those
          memories, and the emotions associated with them.

          What kind of memory would a performer need to draw from
          to imbue words like 'sheep' and 'chain' with sadness?
               One actress…

AUDIOSET: Animal > Livestock, farm animals, working animals > Sheep > Bleat
#2078
          remembers a time when she was a child and her pet sheep
          died. She remembers the sadness she felt, and how her
          parents tried to console her.

Toronto Emotional Speech Dataset YAF_sheep_sad.wav
Toronto Emotional Speech Dataset YAF_death_sad.wav
Toronto Emotional Speech Dataset YAF_pain_sad.wav
Toronto Emotional Speech Dataset YAF_time_sad.wav
Toronto Emotional Speech Dataset YAF_young_sad.wav
Toronto Emotional Speech Dataset YAF_learn_sad.wav
Toronto Emotional Speech Dataset YAF_take_sad.wav
Toronto Emotional Speech Dataset YAF_lose_sad.wav
Toronto Emotional Speech Dataset YAF_far_sad.wav
Toronto Emotional Speech Dataset YAF_whole_sad.wav
Toronto Emotional Speech Dataset YAF_voice_sad.wav

Music made with Bugbrand Board Weevil circuit board. AUDIOSET: Natural sounds
> Water > Ocean > Waves, Surf #2777

                    VOICE 3
          The other actress remembers being at the beach with her
          friends and getting her foot caught in a chain. She re-
          members the pain she felt, and how her friends helped
          her get free.

Toronto Emotional Speech Dataset OAF_chain_fear.wav
Toronto Emotional Speech Dataset OAF_hole_fear.wav
Toronto Emotional Speech Dataset OAF_chain_anger.wav
Toronto Emotional Speech Dataset OAF_hole_anger.wav
Toronto Emotional Speech Dataset OAF_chain_neutral.wav
Toronto Emotional Speech Dataset OAF_deep_fear.wav
Toronto Emotional Speech Dataset OAF_chain_sad.wav
Toronto Emotional Speech Dataset OAF_deep_pleasant_surprise.wav
Toronto Emotional Speech Dataset OAF_kick_fear.wav
Toronto Emotional Speech Dataset OAF_limb_fear.wav
Toronto Emotional Speech Dataset OAF_kick_happy.wav
Toronto Emotional Speech Dataset OAF_limb_disgust.wav
Toronto Emotional Speech Dataset OAF_kick_anger.wav
Toronto Emotional Speech Dataset OAF_beg_fear.wav
Toronto Emotional Speech Dataset OAF_numb_fear.wav

Music made with Bugbrand Board Weevil circuit board

                    VOICE 3
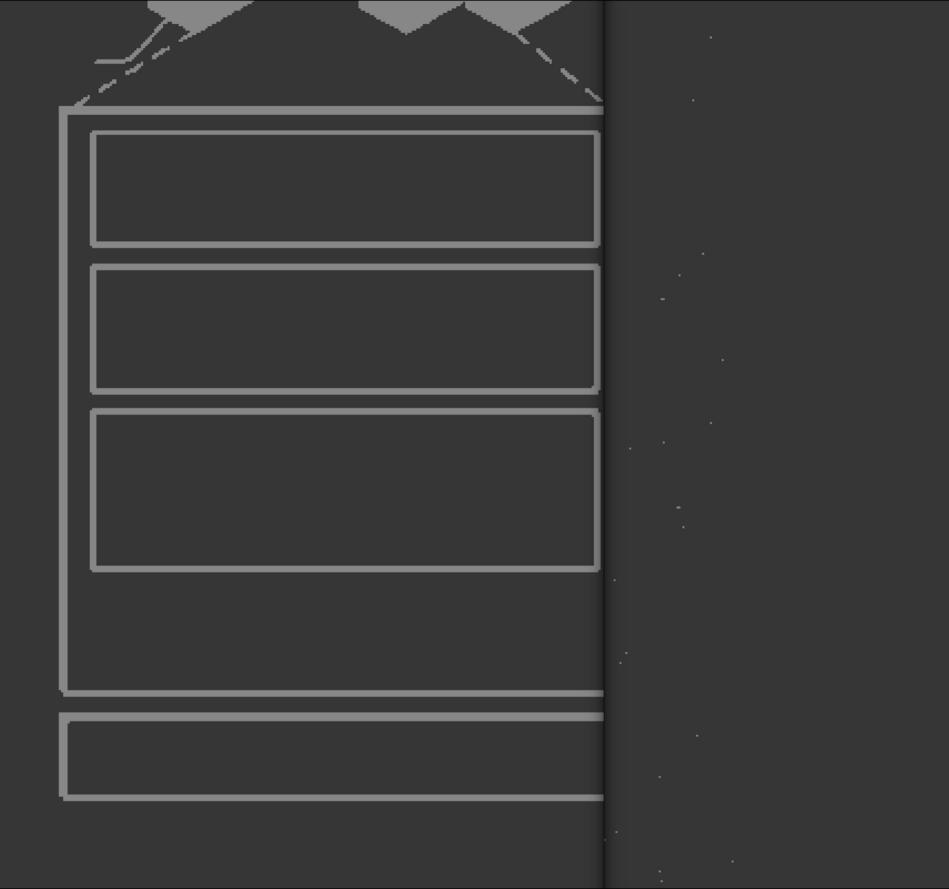   Months later, when the researchers are analysing their
          data, they begin to uncover these embedded memories.

          Should they keep them hidden?

DCASE Sound Event Detection: Office Live Testing Dataset: keys15.wav

          Or share them with the world? Either way, the research-
          ers know that the memories are now a part of the data-
          set, and always will be.

AUDIOSET: Natural sounds > Water > Ocean > Waves, Surf #2777. DCASE Sound
Event Detection: Office Live Testing Dataset: keys15.wav

<u>SCENE 8 — IN THE REAL WORLD</u>

Music made with samples from 19th century French music boxes

                    VOICE 1
          Imagine someone standing on an otherwise empty stage.

Echo, as if in a large empty room

          It's mostly dark with blue and purple tones. It's
          glossy. Big red letters, there's a D, an E, and a T.

Music made with Bugbrand Board Weevil circuit board cuts in

                    VOICE 4
          Everything in the real world is being recreated in the
          virtual world. The metaverse. The metaverse, a per-
          sistent digital universe that mirrors our world, but
          is becoming as diverse and awe-inspiring as the natural
          world. And you'll be able to do everything you do in
          the real world in the virtual world, and more.

          In the real world, you can touch a rock, and in the
          metaverse, you can touch a rock. You can pick it up,
          you can throw it, you can break it. You can interact
          with it in ways that are impossible in the real world.

          In the real world, you can buy a house, and in the
          metaverse, you can buy a house. But in the metaverse,
          you can also buy a moon, a sun, or a star. You can buy
          anything you can imagine, and more.

          In the real world you can say a word. But in the
          metaverse you can own every word you speak. Or pay rent
          to the owner or maybe buy the licensing rights to the
          word and give you a steady income stream to support
          your use of other people's word. The possibilities are
          endless.

          I make 20 words every day, just in case. And you can
          too.
                    VOICE 1
          Say the word DEAKSPOOK.

                    VOICE 3
          DEAKSPOOK.

                    VOICE 1
          Say the word SOCKEDGEND.

                    VOICE 6
          SOCKEDGEND.

                    VOICE 1
          Say the word WAKE.

                    VOICE 2
          WAKE.

                    VOICE 1
          Say the word WORD.

                    VOICE 4
          WORD.

                    VOICE 1
          Say the word SARCASTICALLY.

                    VOICE 3
          SARCASTICALLY.

                    VOICE 1
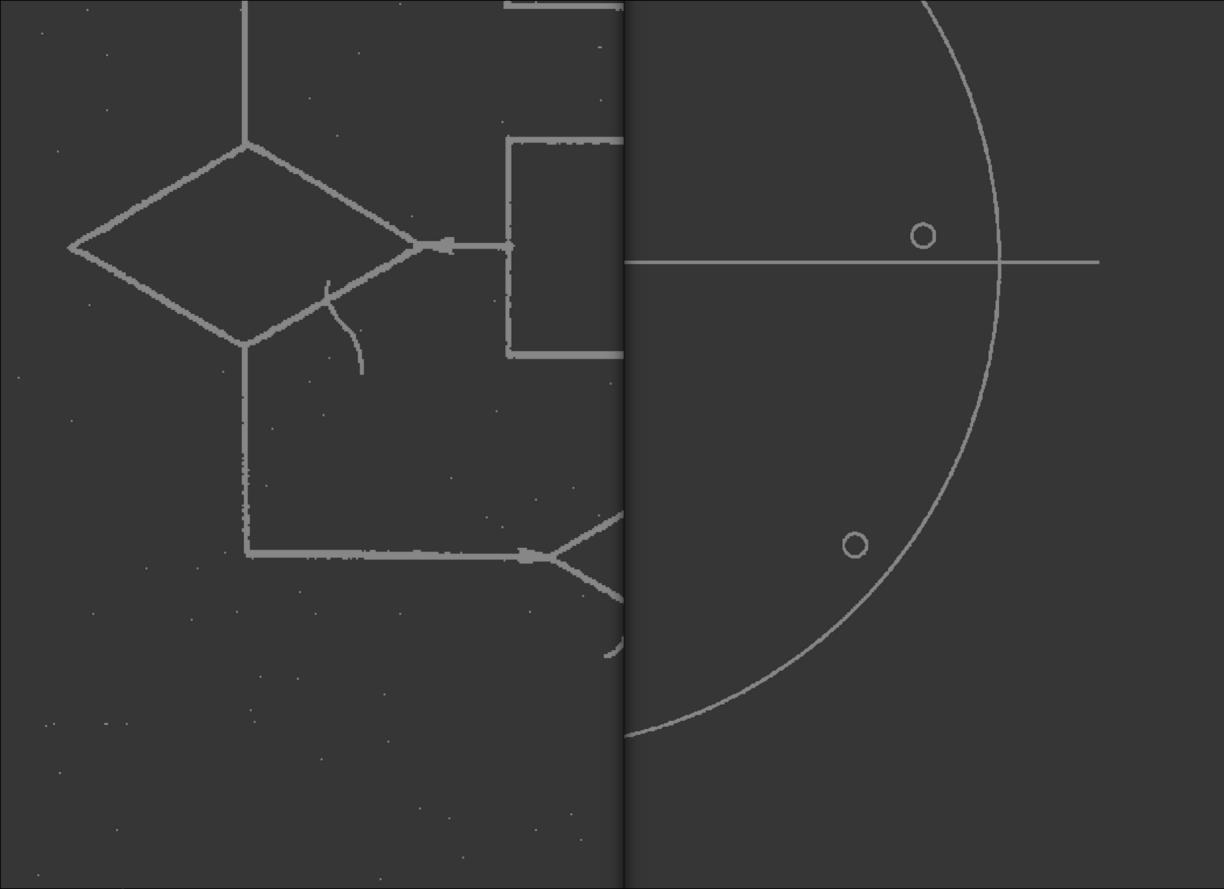   Say the word VOICE.

                    VOICE 6
          Voice.

                    VOICE 1
          Say the word AHHHH.

Sustained vowels from Consensus Auditory-Perceptual Evaluation of Voice Data-
set

                    <u>LOOP TO START</u>