

Consent & Eligibility

Hello! To answer this survey, you should be a viewer of video content. If you are a video captioner or creator, [you can also answer the captioner survey here](#). In this section, we will go through your survey consent, age eligibility, and prizes for completion.

Consent Form for IRB-FY2023-7683

Good day! You have been invited to take part in a research study to learn more about the accessibility of non-speech sound captions. This study will be conducted by Mark Cartwright, STEINHARDT - Music and Audio Research Laboratory (MARL), Steinhardt School of Culture, Education, and Human Development, New York University.

If you agree to be in this study, you will be asked to take part in an online survey about experiences with non-speech sound captioning and demographic questions related to age, race, gender orientation, etc.

Participation in this study will involve 30 minutes to complete the survey. There are no known risks associated with your participation in this research beyond those of everyday life.

Although you will receive no direct benefits, this research may help the investigator understand the accessibility of non-speech sound captions. Participants in the survey will be entered into a raffle to receive one of 10 Amazon gift cards (\$40 each).

In the survey, you will be offered to participate in an interview. If you choose to participate in the interview, personally identifiable data from the survey will be linked to the interview data. If you choose not to participate in the

interview, confidentiality of your research records will be strictly maintained by assigning code numbers to each participant so that data is never directly linked to individual identity. Also, collected data will be stored on the NYU Google Drive storage with two-factor authentication and researchers' password-protected computers and only accessible to the research personnel. Information not containing identifiers may be used in future research, shared with other researchers, or placed in a data repository without your additional consent.

Participation in this study is voluntary. You may refuse to participate or withdraw at any time without penalty. For interviews, questionnaires, or surveys, you have the right to skip or not answer any questions you prefer not to answer.

If there is anything about the study or your participation that is unclear or that you do not understand, if you have questions or wish to report a research-related problem, you may contact Mark Cartwright at 773-850-1527, mc6591@nyu.edu, 370 Jay Street, 3rd Floor, Brooklyn, NY, 11201.

For questions about your rights as a research participant, you may contact the University Committee on Activities Involving Human Subjects (UCAIHS), New York University, 665 Broadway, Suite 804, New York, New York, 10012, at ask.humansubjects@nyu.edu or (212) 998-4808. Please reference the study # (IRB-FY2023-7683) when contacting the IRB (UCAIHS).

You have received a copy of this consent document to keep.

Agreement to Participate

I agree to the above and participate in this survey.

Completing the survey will enter you in a raffle for a \$40 Amazon gift card. Participation in a follow up interview will guarantee you a \$30 Amazon gift

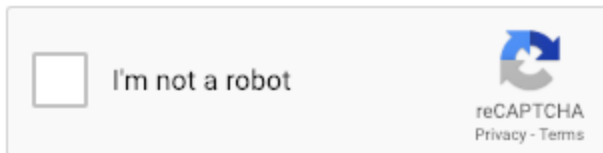
card. Please enter your contact details at the end of the survey so that we can contact you about these rewards.

I understand that at the end of the survey, I can fill in my contact details in order to receive rewards.

Are you at least 18 years old and either fluent in English or American Sign Language?

Yes and continue.

Please complete the captcha verification.



Demographics

How old are you?

Choose one or more races that you consider yourself to be

White or Caucasian

Black or African American

American Indian/Native American or Alaska Native

Asian

Native Hawaiian or Other Pacific Islander

Other

Prefer not to say

How do you describe yourself?

Male

Female

Non-binary / third gender

Prefer to self-describe

Prefer not to say

What is the highest level of education you have completed?

Some high school or less

High school diploma or GED

Some college, but no degree

Associates or technical degree

Bachelor's degree

Graduate or professional degree (MA, MS, MBA, PhD, JD, MD, DDS etc.)

Prefer not to say

Hearing & Sign Language

Which of these best describes your relationship to hearing? (Tick all that apply)

deaf

Hard-of-Hearing

Deaf

Hearing

Other

Do you use any hearing assistive technologies such as hearing aids or cochlear implants?

Yes

No

What hearing assistive technologies do you regularly use on your **LEFT** side?

Hearing aid

Cochlear implant

Nothing

Other

How long have you been using this hearing technology on your **LEFT** side?

What hearing assistive technologies do you regularly use on your **RIGHT** side?

Hearing aid

Cochlear implant

Nothing

Other

How long have you been using this hearing technology on your **RIGHT** side?

You mentioned that "Deaf" best describes your relationship to hearing. How long have you identified as "Deaf"?

Since birth

Since childhood

Since I was a teenager

Recently, within the past 5 years

Other

You mentioned that "deaf" best describes your relationship to hearing. How long have you identified as "deaf"?

Since birth

Since childhood

Since I was a teenager

Recently, within the past 5 years

Other

You mentioned that "Hard-of-Hearing" best describes your relationship to hearing. How long have you identified as "Hard-of-Hearing"?

Since birth

Since childhood

Since I was a teenager

Recently, within the past 5 years

Other

You mentioned that "Hearing" best describes your relationship to hearing. How long have you identified as "Hearing"?

Since birth

Since childhood

Since I was a teenager

Recently, within the past 5 years

Other

You mentioned that $\{q://QID197/ChoiceTextEntryValue/7\}$ best describes your relationship to hearing. How long have you identified as

#{q://QID197/ChoiceTextEntryValue/7}?

Since birth

Since childhood

Since I was a teenager

Recently, within the past 5 years

Other

How often do you communicate using a signed language, such as ASL or SASL?

A few times a month, I sign occasionally

A few times a week, I sign time to time

Daily, I sign a lot

I do not communicate in a signed language

Other

Which signed languages do you communicate in?

American Sign Language (ASL)

South African Sign Language (SASL)

Signed Exact English(SEE)

Pigeon Signed English (PSE)

Other

How long have you used a signed language to communicate?

Less than 1 year

1-3 years

3 - 5 years

5 - 10 years

10+ Years

Other

Watching Video Content

You will now be asked about video content and closed captioning. **Video content** here is only **pre-recorded (not live) recordings**. **Closed captioning** is when the audio portion of a video is showed as text on the screen.

In a typical **month**, on which of the following platforms do you usually watch **1 hour or more** of video content on? (Please tick all that apply)

Netflix

Max (Formerly HBO Max)

Amazon Prime

Hulu

Disney +

Cable TV or DSTV

YouTube Shorts

YouTube (Regular)

TikTok

Instagram Reels

ESPN+

Other

In a typical month, which of the following **genres** of video content do you usually watch more than 1 hour of? (Please tick all that apply)

Action

Sci-Fi

Drama

Sport

Music videos or musicals

Comedy

News

Educational or instructional

Documentary

Travel

Vlogs

Reviews

Fashion

Fitness

Food/Cooking

Animation

Kids

Other

In a typical **day**, how many hours of video content do you watch? This includes streaming, videos on social media & YouTube, and regular TV.

0-1 Hours

1-2 hours

2-3 Hours

3-4 Hours

4+ Hours

How often do you use **closed captioning** when watching videos? When they are available, I use closed captions...

Rarely

Occasionally

Most of the time

Always

I never use closed captions

General_NSI

You will now be asked about **captioning for non-speech information**. Non-speech information (NSI) is extra information about sounds and dialogue besides the words spoken. Some examples include music, footsteps, or the emotion with which someone says something.

How important is it to you that **non speech information (NSI)** be captioned?

Not at all important

Slightly important

Moderately important

Very important

Extremely important

Why do you think **non speech information (NSI)** should be captioned?
(Tick all that you agree with)

It adds to viewer enjoyment

It is important in understanding the video content

It is legally required

It should not be captioned

Other

In general, how satisfied are you with captions for **non-speech information (NSI)**?

Extremely dissatisfied

Somewhat dissatisfied

Neither satisfied nor dissatisfied

Somewhat satisfied

Extremely satisfied

Please rank the following issues/problems with captions for **non-speech information (NSI)**, with 1 being the biggest problem and 9 being the smallest problem

The NSI seems redundant with the visuals

Words used to describe the NSI are confusing

Other

NSI that is present contains too much detail (i.e., captions are too long to read)

The NSI seems irrelevant

NSI that is present lacks detail

The NSI revealed hidden narrative information (e.g. a character's name or the source of the sound)

Lack of captions for NSI in general

Timing of the NSI captions seems off

When **NSI** is NOT captioned, do you have other ways to get that information? Tick all that apply.

☐ Using an app on an external device (list the app here:)

Asking friends or family who are watching with me

Text or send a video to a friend or family member

Rewind the video

☐ Other

I do not use any other ways to get that information

Are there types of NSI that are generally **not** captioned? If yes, please list the NSI you would like captioned but is generally not captioned.

Are there any other issues you experience with **captions for NSI**?

How frequently does this occur to you while you are watching videos?

	Never	Rarely	Sometimes	Very Often	Always
I am confused who is speaking	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The tone or emotion of a scene is unclear to me	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The tone or emotion of a character who is speaking is unclear to me	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The timing of things like punch line of jokes, or jump scares in horror movies feels a bit off	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am confused about an off-screen event (something that is happening in the background)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Speech nuances and hidden meanings can be confusing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Never	Rarely	Sometimes	Very Often	Always
I react out-of-sync with other audience members who are not using captions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

NSI_SFX

You will now be asked questions about captioning **non-speech sound effects (NSI sound effects)**. These are sounds like footsteps, laughing, gunshots, etc that might be captioned like "[door creaks]" or "[footsteps in the distance]", as well as background sounds.

I think **NSI sound effects** for video content are captioned:

	Strongly Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Strongly Agree
Often enough	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
With enough detail	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
With too much or irrelevant detail	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
With inaccurate information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Are there video platforms or genres where issues with captioning for **NSI sound effects** is good or bad? If so, please write the name and/or genre (e.g. "Action movies on Netflix", "News on CNN")

Please rank the importance of these characteristics of **NSI sound effects**. (1 being most important and 8 being least important)

The exact timing of the sound (e.g. "Gunshot" at the exact moment the gun fires)

The type of sound (e.g. "Gunshot")

The timbre or "color" of the sound (e.g. "delicate", "harsh")

The action that caused the change in sound? (e.g. "turns off faucet" rather than "dripping water stops")

The volume of the sound (e.g. "loud", "quiet")

The beat or rhythm of a sound (e.g. "Fast pulsating")

Where the sound is coming from (e.g. "from the left", "behind the fridge")

The mood or emotion being conveyed (e.g. "joyous", "mournful")

Frequency information (e.g. "high-pitch", "low-rumble")

How much **NSI sound effects** do you think should be communicated?

Only non-obvious sounds important to the video content

Sounds important to the video content AND sounds that establish the tone or emotion

Most sounds

Every sound

I am not concerned with the details

Other

I want **NSI sound effects** to be described by...

Brief, factual descriptors e.g. ("Door Slams", "Foot steps", "Music")

Some additional factual details e.g. ("Door Slams Loudly", "Quick Foot Steps", "Up-Tempo Music")

Some additional affective details e.g. ("Door Slams Angrily", "Scared Foot Steps")

Some combination of additional factual and affective details e.g. ("Door Slams Loud and Angrily", "Quick, Scared Foot Steps")

Mainly factual details e.g. ("Door Slams Loudly Behind The Couch with a Slight Squeak", or "Quick Foot Steps of High-Heel Shoes on a Tiled Floor")

Mainly affective details e.g. ("Door Slams Furiously, Creaking with Anger", or "Scared Foot Steps Lightly Hurry Away")

No preference

Music

You will now be asked questions about captions for **music**. This includes background music as well as music made by actors or people in the video.

Please rate how much you agree with the following: When I watch videos, I am often unsure about...

	Strongly Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Strongly Agree
The tone or emotion of music	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The level of importance of music. (Is it just background music or is it important to the story)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Whether the same piece of music is being played multiple times throughout a video	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Are there platforms or genres where captioning for **music** is particularly good or bad? If so, please write the name and/or genre (e.g. "Action movies on Netflix", "News on CNN")

Please rank which qualities of **music** you want to know while watching videos. 1 is the most important, and 9 is the least.

The style or genre of the music (e.g. "Jazz", "Rock")

Frequency information (e.g. "high-pitch", "low-rumble")

The mood or emotion being conveyed (e.g. "joyous", "mournful")

The timbre or "color" of the sound (e.g. "delicate", "harsh")

The timing of the music start and stop

The beat or rhythm in the music

Where the music is coming from (e.g. "from the left", "from the radio")

The volume of the music (e.g. "loud", "quiet")

Whether the music originates from within the scene (diegetic - e.g. music in an elevator vs non-diegetic - e.g. narration)

Which amount of **music** do you want communicated in videos?

I only want music that is important to the video content communicated to me

I want all music that is important to the video content as well as music that establish the tone or emotion communicated

I want most music communicated to me

I want all music communicated to me

I am not concerned about music

Which details of **music** do you want communicated to you?

I only want brief, factual descriptors of the music style e.g. "Jazz Music", "Rock Music"

I want some additional factual details e.g. "Loud Jazz Music", "Slow Rock Music"

I want some additional emotional details e.g. "Uplifting Music", "Depressed Music"

I want details about the music name and artist e.g. "Yesterday - The Beatles"

I want details about the genre, name, and artist e.g. "Yesterday - The Beatles (Light Rock Music)"

I want a lot of emotional details e.g. "Bright, cheery music with an uplifting spirit", "Depressed music dragging like a bag of bricks"

Other

Speaker Manner of Speech (Speech Delivery)

You will now be asked about captions for **Manner of Speech**. This includes things like if someone is shouting, whispering, saying something with an accent, sarcastically, etc.

Please rate how much you agree with the following:

	Strongly Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Strongly Agree
I am often unsure of the manner of speech someone is using	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When characters are not facing the camera, I am often unsure of the manner of speech someone is using	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When I am unsure of the manner of speech, it is difficult for me to understand the video content	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Are there video platforms or genres where captions for **manner of speech** are very good or very bad? If so, please write the name and/or genre (e.g. "Action movies on Netflix", "News on CNN")

How frequently do you want **manner of speech** to be communicated?

I never want manner of speech communicated

I only want manner of speech communicated when it is either narratively important or not obvious from the video itself

I want manner of speech communicated most of the time

I always want manner of speech communicated

Which qualities of **Manner of Speech** would you like captioned? 1 is most important and 4 is least important

Other non-speech sounds e.g. "Laughing", "Coughing"

Intentional deviation from normal speech e.g. "Fake British Accent", "Sarcastically"

Volume e.g. "Shouting", "Whispering"

Emotion or affect e.g. "Angrily", "Excited"

How much details of **manner of speech** do you want captioned?

Brief descriptors e.g. "Whispering", "Shouting"

Descriptors with extra factual details e.g. "Gently Whispering", "Shouting Loudly"

Descriptors with lots of details e.g. "Gently Whispering like a Breeze", "Earth-shakingly Loud Shouting"

I am not particularly concerned about these details

Speaker Label

You will now be asked questions about captions for **Speaker Identification / Speaker Labels**. This includes ways to communicate which person spoke which words.

How much do you agree with the following statements about **Speaker Labels**?

	Strongly Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Strongly Agree
When there are multiple characters on screen, I am often confused who is speaking	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When characters are not facing the camera, I am often confused who is speaking	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When I am unsure of who is speaking, it is difficult for me to understand the video content	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is frustrating when a character's name appears in captions before it is revealed in the video content itself	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Are there video platforms or genres where captions for **speaker labels** are very good or very bad? If so, please write the name and/or genre (e.g. "Action movies on Netflix", "News on CNN")

What frequency do you want **speaker identification** communicated to you?

I never want any speakers identified

I only want speakers identified if it is not clear in the video who is speaking

I want speakers to be identified whenever they first start talking

I always want speakers to be identified

Please rank the following ways to communicate **speaker labels**. 1 is the way

you prefer most and 8 is the way you prefer least:

1 2 3 4 5 6 7 8

Color, Name, & Alignment



Color & Alignment



Color



Aligned



No speaker identification



1 2 3 4 5 6 7 8

Color & Name



○○○○○○○○

Name & Alignment



○○○○○○○○

Name



○○○○○○○○

Novel Technology

You'll now be shown pictures of 5 technologies designed to help communicate non-speech information and be asked to answer a few questions about each of them.

T1

Text Modulation: This is when the font, color, or style of closed captions is modified to show a speaker's voice, pitch, loudness, or emotion.

Text Modulation



For which of the following **genres** of video content would this technology be useful?

Action

Sci-Fi

Drama

Sport

Music videos or musicals

Comedy

News

Educational or instructional

Documentary

Travel

Vlogs

Reviews

Fashion

Fitness

Food/Cooking

Animation

Kids

Other

Any thoughts or comments about this technology?

T2

Comic Style Captions: This is when the closed captions are placed around characters in the way comic books place dialogue

Comic Style Captions



For which of the following **genres** of video content would this technology be useful?

- Action
- Sci-Fi
- Drama
- Sport
- Music videos or musicals
- Comedy
- News
- Educational or instructional
- Documentary
- Travel
- Vlogs
- Reviews
- Fashion

Fitness

Food/Cooking

Animation

Kids

Other

Any thoughts or comments about this technology?

T3

Emoji or Icon-Supported Captions: This is icons or emojis are used to express speaker emotions in closed captions.

Emoji or Icon Supported Captions



For which of the following **genres** of video content would this technology be useful?

Action

Sci-Fi

Drama

Sport

Music videos or musicals

Comedy

News

Educational or instructional

Documentary

Travel

Vlogs

Reviews

Fashion

Fitness

Food/Cooking

Animation

Kids

Other

Any thoughts or comments about this technology?

T4

Customizable Captions: You are able to choose which characters are captioned, as well as the colors associated with them

Customization

Caption Customization Settings



Rachel McAdams

☒ Captions

Color: Yellow



Ryan Gosling

☒ Captions

Color: Green



For which of the following **genres** of video content would this technology be useful?

Action

Sci-Fi

Drama

Sport

Music videos or musicals

Comedy

News

Educational or instructional

Documentary

Travel

Vlogs

Reviews

Fashion

Fitness

Food/Cooking

Animation

Kids

Other

Any thoughts or comments about this technology?

T5

Sound Visualization: This would show aspects of audio through an audio-reactive animation, such as showing where sounds are coming from and how loud sounds are using animated lines.

Sound Visualization



For which of the following **genres** of video content would this technology be useful?

- Action
- Sci-Fi
- Drama
- Sport
- Music videos or musicals
- Comedy
- News
- Educational or instructional
- Documentary
- Travel
- Vlogs
- Reviews
- Fashion

Fitness

Food/Cooking

Animation

Kids

Other

Any thoughts or comments about this technology?

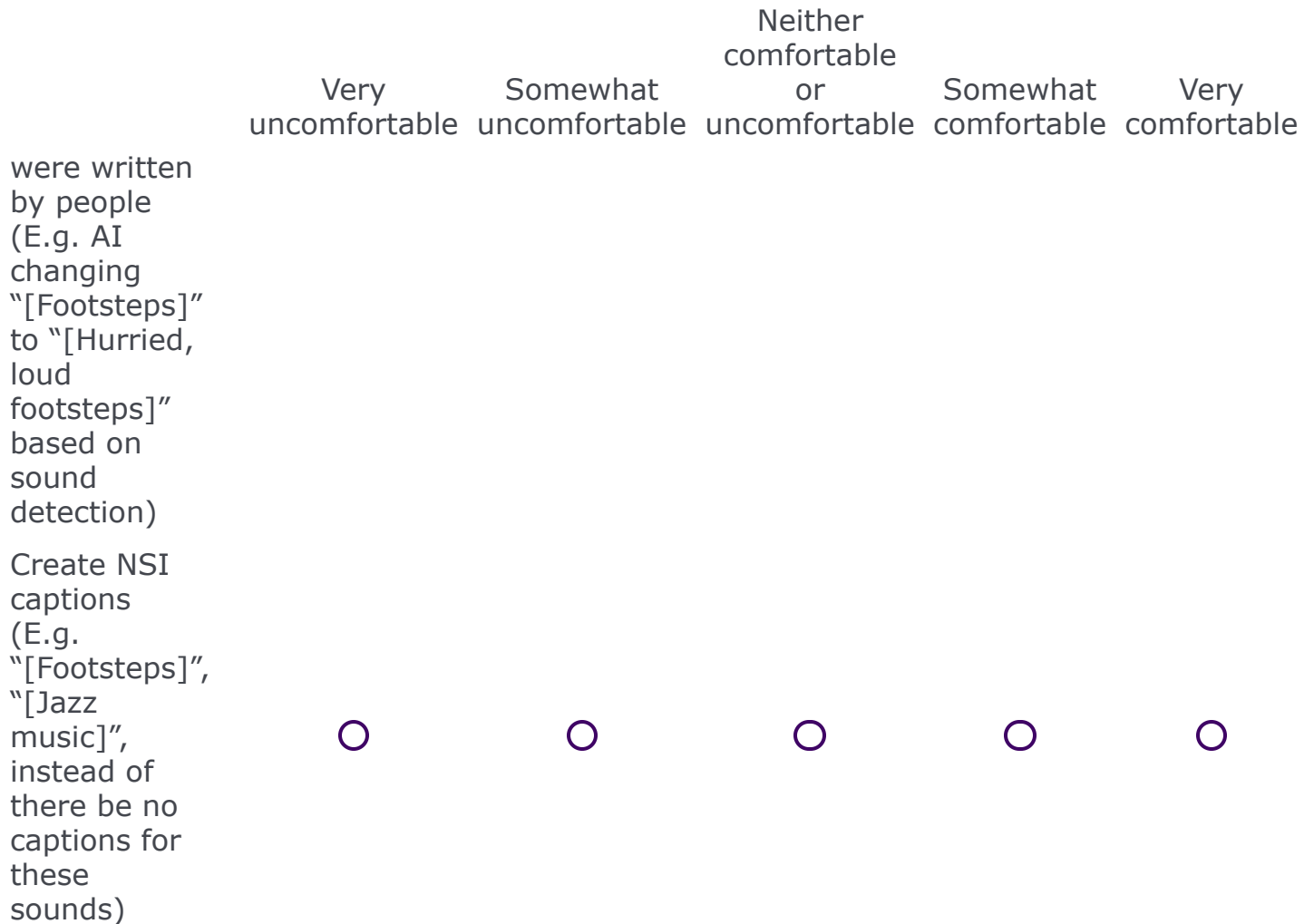
Novel Tech After

You've seen 5 examples for current captioning technologies in development. What technology to visually communicate information about sound in a video do you wish existed?

AI Captions

Please rate how comfortable you would feel about artificial intelligence (AI) being used to:

	Very uncomfortable	Somewhat uncomfortable	Neither comfortable or uncomfortable	Somewhat comfortable	Very comfortable
Modify NSI captions that	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Checking Interview Participation

We are almost there! Please complete a few more questions to be entered into the raffle for the \$40 Amazon Gift cards.

Please create or write your participant ID. This would be the two digits of the day of the month you were born and the last 2 digits of your cell phone number. (E.g. 15 December birthday, Phone Number 999-999-999 would be 1599)

If you would like to join the survey raffle for \$40 Amazon gift cards, participate in an interview for \$30 Amazon gift card, or be notified about future research, [please fill out this 1-minute form for your contact details](#).

I have filled in the survey with my details.

I do not want to participate in the raffle, interview, or future research.

Powered by Qualtrics