Secondary Encoding

[authorship removed for blind review]

Translation between media is often used to make information encoded in one modality available to users who are unable to make use of that modality. Translation acts on the primary means of encoding the information (e.g., text, sound). Secondary encoding (e.g., color, intonations) is often added to primarily encoded information to help the user correctly interpret its intended meaning. During translation between media, secondary encodings are often removed and their information remains inaccessible. When this happens, miscommunications may occur. Although secondary encoding is a multi-discipline problem, different types of secondary encoding are usually analyzed as separate problems. In reality, they are aspects of a single problem. This paper presents a taxonomy of secondary encodings and guidance for improving the accessibility of secondary encoded information.

INTRODUCTION

When people communicate, information may be exchanged (both consciously and unconsciously) using multiple types of encodings and media. Examples of encodings include speech, and body language. Multiple encodings can complement each other, providing more information as a whole. Because encodings can be presented together, it may be difficult to capture the full meaning when any encodings disappear.

Accessibility requires all information be perceived either directly or in some form of translation. With assistive technologies, information from various encodings can be translated between modalities. The problem is that many current translation solutions leave out significant amounts of information, thus the information becomes inaccessible. Additionally, people who are capable of using one modality for information may have certain disabilities that cause them to miss some information provided subtly via the same or another modality.

This paper is concerned with two types of encoding, referred to as primary and secondary encoding. While primary encoding is more likely to be made accessible, secondary encoding is often overlooked and is seldom made accessible.

For the purposes of this paper, primary encoding is defined as the "astylistic" presentation of information via a medium of communication such as speech, text, and hand gestures. It is the actual words or actions presented. From the paralinguistic point of view, primary encoding is "what" is presented (Campbell, 2005).

One primary encoding can complement another primary encoding. For example, two primary encodings in a movie are sound and video. Someone with a visual impairment may only receive the audio information. To truly understand the entire movie, the person needs to know the visual information as well. Descriptive audio attempts to provide the person with some of the missing visual information. However, some information (e.g. information on the body language of the speaker) is more subtle than is found in traditional descriptive audio.

For the purpose of this paper, secondary encoding is defined as a presentation style that applies to or accompanies primary encoding and provides additional information. An example of the application of secondary encoding is a significant color being applied to visually displayed text. An example of

the accompanying of secondary encoding is the use of gestures to emphasize speech. From the paralinguistic point of view, secondary encoding is "how" primary encoding is presented or emphasized (Campbell, 2005).

Secondary encoding is often used to convey information, whether people realize they are doing it or not. The way the information is conveyed may not be detected by all people and could be misinterpreted by those who detect it. When the secondary encoding is removed or lost, a portion of the information goes with it, which can lead to further miscommunication. Some information may become lost when the primary encoded information is translated to another medium without consideration of accompanying secondary information.

Instances of secondary encoding (as discussed below) have been the subject of years of research; however, no generalized concept has yet been developed. While secondary encoding is a cross-discipline problem, each discipline uses different terminology for aspects of the same problem. Each variation of secondary encoding is generally considered separately. As a result, there is more information on certain problems than others. Attempts have been made to resolve some of the identified problems, but not all problems.

This paper attempts to gather and present the various aspects in a taxonomy of secondary encoding techniques.

BACKGROUND

This section describes secondary encodings from different disciplines, followed by how the pairings of encodings can affect accessibility. Finally, a set of accessibility issues are presented.

Multi-Discipline Concept

Since the use of secondary encoding is pervasive, aspects of secondary encoding are studied in many disciplines.

Linguists use the term "paralanguage" or "paralinguistic". According to Poyatos (1993), paralanguage is "the nonverbal voice qualities, modifiers and independent sounds…". Paralanguage focuses on eight primary qualities of speech, including loudness, tempo, and intonation range. These qualities can be considered secondary encodings to the primary encoding of

speech. Paralanguage focuses on speech and ignores the encodings of the other media. Psychologists tend to refer to similar concepts as "non-verbal communication" (Harrison, 1974).

Ongoing multimedia research is attempting to find means of including emotions and other forms of nonverbal communication (secondary encodings) in various technologies (Oudeyer, 2003).While many people are researching this issue, they are not dealing with it as an accessibility problem.

Although few complete solutions exist, research has been done to make some secondary encoded information accessible. The Centre for Learning Technologies group at Ryerson University has developed technology for inserting emotions and sound into videos for the hearing impaired (Fels et al, 2005).

Color has been considered as a problem on its own and many solutions have been developed for it. The World Wide Web Consortium (W3C, 1999) Web Content Accessibility Guidelines (WCAG) has a guideline regarding its use. Few guidelines exist for the other types of secondary encoding.

Synergy and Cognitive Dissonance

Gestalt theorists have a central idea that the whole is greater than the sum of its parts. All of the separate parts can provide more information when presented together. There is very little information available when only one part is present.

When the encodings are separated, the meaning that each provides has less impact individually. When the secondary encodings are removed from the primary encoding, the only information that is left is in the primary encoding. The primary encoding alone does not provide as meaningful information.

Sometimes, the encodings provide contradictory information, resulting in confusion and create cognitive dissonance. Festinger (1966) defines cognitive as "any knowledge, opinion, or belief about the environment, about oneself, or about one's behavior." and dissonance as "...the existence of nonfitting relations among cognitions..." This causes accessibility problems because if different secondary encodings present conflicting information, then those who cannot detect all of the secondary encodings would not know that the information they receive may be inaccurate.

For example, in the Stroop test, the participant is to state the color of the word as fast as possible. If the word "red" is written in green, the correct answer is "green". It would take more time to arrive at the answer since the text and color of the text provide conflicting information, resulting in "motivational tension" (Wicklund & Brehm, 1976). A person who can only see black and white would provide the incorrect answer of "red" since they cannot detect the color of the word.

Accessibility Issues

Secondary encodings can prove to be challenging to a large portion of the population (due to sensory, perceptual, cognitive, physical, and other issues). Some people cannot or may not perceive secondary encodings when they are presented in a modality other than the one used for primary encoding. For example, someone with who is blind can easily hear speech but will not perceive information from body language. This affects all people with total or partial impairments in a certain modality.

Some people may not detect the secondary encoding in a specific modality even though they can perceive information in that modality. They are imperceptive of non-verbal information due to cultural differences and various social and/or cognitive disabilities. An example of this is some cultures use specific tones in their speech. A person may not be able to detect tones that are not used in their culture.

Some people may not realize that secondary encoding was ignored or missed during translation between modalities. It can be translated back into the original modality but the missing secondary encoding cannot be recovered. For example, synthesized speech cannot add the correct emphasis without explicit instructions because it does not understand the context of the content. Thus, someone who is capable of detecting emotions in speech does not receive this information when synthesized speech is used to provide the information.

Because people are affected by secondary encoding, the removal of it can affect everyone. Secondary encodings may be lost during translation or due to the inability of an individual to detect it. Information that is available only via secondary encoding needs to be presented in a modality that the individual can perceive.

MAKING SECONDARY ENCODING INFORMATION ACCESSIBLE

Information from secondary encodings can be important, therefore, it is essential that the information be made accessible. A person can be fully informed when they possess all of the information. To accomplish this, the following design process is proposed.

1. Identify the presentation styles that are used for secondary encoding. It is important to know the different ways that secondary encoded information is presented.
2. Define the information provided by the specific secondary encoding. The information presented must be concretely identified in order to be expressed in a different manner or modality.
3. Re-encode the information into a modality that is accessible. This step inserts the secondary encoding information into a modality that is accessible to the receiver. As a result, the modality retains all of the original information. The modality may or may not use the same modality as was originally presented.

TAXONOMY OF SECONDARY ENCODING TECHNIQUES

There are three different modalities through which secondary encodings can be portrayed by current information technologies: hearing (auditory), seeing (visual), and touching (tactile). These will be considered to be the possible channels that secondary encodings can be detected and conveyed.

There are essentially two means by which secondary encodings can be presented: they can be added as an attribute of

the primary encoding medium or they can be added via other media. For example, volume and body language are both secondary encodings of the primary encoding speech. Volume is an attribute of speech, while body language uses the visual medium to present the information.

Auditory Secondary Encodings

This section describes four types of secondary encoding within the auditory modality: intonation, stress, volume, and speed.

Intonation. The tone of a person's speech conveys information about their emotional and mental state at that moment. A single phrase can portray different information depending on tone used. For example, "I'm fine." could mean the speaker is actually fine or the speaker is annoyed and wishes to be left alone. It is difficult to determine the meaning when the only information available is the text.

Each person has a certain range of sound that they can detect. Some people can hear sounds at only high pitches, others at only low pitches, and some people have gaps in their hearing range. Information provided outside of the listener's range of hearing becomes missing and the listener may not know that it exists. Auditory alerts (earcons) presented outside of the listener's range of hearing do not achieve their purpose of informing the listener of events.

Stress. Like intonation, stressing or emphasizing different words within a sentence can assign a different meaning to a sentence. For example, in "She's going.", stressing "She's" would mean she is the only one going and no one else. Stressing "going" would mean she's not staying.

Those with the ability to speechread may be unable to detect oral emphasis because the words are produced using the same movements of the mouth. Both the emphasis and the meaning of the words may be missed in the new medium when a sentence is translated from speech to another medium.

Volume. Volume can be a deceptive attribute within speech. The volume with which a person speaks can signify certain emotions. Sobin & Alpert (1999) have associated increased volume with anger and decreased volume with sadness. However, this is not always the case. A person sounds louder when they are near and quieter when they are distant. Sometimes, people speak loudly with the misconception that they will be heard more easily. At other times, the person naturally speaks at a volume due to their culture.

The volume of a person's speech can affect both the primary and secondary encoding. The information would not be perceived if the speech is at a volume that the listener cannot detect. For example, a person may miss emotional and cultural information as they strain to hear the words (recognizing the primary encoding) being spoken.

Speed. Speed can indicate a person's emotional state at that moment in time. When a person is nervous, their speech may speed up. When a person is calm, their speech is constant and the words are deliberate.

The speed of speech affects the impact that it makes on the audience. Speaking too fast allows little time for the listener to hear and process the information. Speaking too slow can bore the listeners. An accessibility problem arises since not everyone will receive all of the information.

Pauses affect the speed of speech and can have different effects depending on their usage. When people are nervous, they pause because they do not know what to say. However, when used appropriately, pauses add emphasis and have a great impact. For example, Martin Luther King, Jr.'s legendary "I Have a Dream" speech made effective use of pauses.

Visual Secondary Encodings

This section describes three types of secondary encoding within the visual modality: color and contrast, structure and style formats, and body language.

Color and contrast. Color is often used in documents to add emphasis or additional meanings to certain words. Those words stand out from the rest of the document and attain the reader's attention. This can become a problem for those with color blindness because they cannot differentiate that word from the rest. For example, words written in yellow may not be detected by someone with tritanopia, a blue-yellow deficiency.

Colors in a document also affect those without color blindness. Many websites and advertisements use color to increase attractiveness. However, if chosen incorrectly, the colors can decrease the readability of the content and make it inaccessible. For example, using a light yellow font on a white background is not as effective as using a light yellow font on a black background. Contrasting colors are important because they improve the readability of the content, and become more accessible to more people.

In addition to color differences, designers should consider the differences produced when images are transformed into their grayscale equivalents (which are often produced when printing in black and white). Ensuring grayscale contrasts and differentiation can help in situations where a distinction cannot be made between colors. However, it is more difficult for users to attach significant meanings to different shades of gray similar to the way that meanings are often attached to colors.

Structure and style formats. The structure and headings of a document provide the reader with a quick overview of the document's contents and purpose. For example, a textbook has one column and headings of different styles while a newspaper article may have more columns and no headings. People using screen readers are unable to quickly get an overview of the contents without suitably marked-up headings. If the mark-up does not follow standard conventions, graphically distinctive headings will not be readily available to screen readers to provide navigation support to blind users.

Like stress in speech, the style of text (bold, italics, underline, and font size) also emphasizes certain words. This information may not be conveyed to those with text disabilities. Screen readers may not have the ability to inform the user that the font style has changed. If the document does not use tags the screen reader recognizes (e.g. <STRONG> instead of <B> for bold), the emphasized words could lose their importance (Pennsylvania State University, 2005).

Body language. Body language combines facial expressions and gestures to convey a singular message. When a person communicates, the whole body is used to re-enforce a message. For example, an anxious person may avoid eye contact and start fidgeting. Without the information from body language, the visually impaired cannot judge a person's reaction or comfort level.

Each facial expression that a person makes can be associated with a certain emotion. Ekman and Oster (1979) located studies that found facial expressions for a number of emotions, including disgust, surprise, and distress.

Online conversations are stripped of facial expressions since it is mainly done through text. Even though a sighted person can normally detect facial expressions, this information is not available in text-only communications. Emoticons attempt to insert facial expressions into text. However, the set of commonly recognized emoticons for facial expressions is still limited and there is no definitive legend for their meanings.

Even though alternative-format descriptions of facial expressions could be made available to those with visual impairments, they seldom are. Possibly more important is the meaning being conveyed by those expressions. However, this meaning is also seldom available.

People often use gestures with their hands as they speak. The use of hand gestures is often culturally dependent and can be difficult to control. The gestures serve as visual cues. The wider the hand gestures made by a speaker, the stronger the emotion and the wilder the hand gestures, the greater the excitement. The speed of the gestures also acts as a cue regarding the intensity of the emotion. Likewise, the conductor of an orchestra uses the speed and extent of movement, to direct the speed and volume of the music. With signed languages, the speed and the stress of the signs provide similar secondary encodings.

Without the ability to see the gestures and cues, the visually impaired may not receive the secondary (cultural and emotional) information that is being expressed. It is important to possess this information because very few other secondary encodings demonstrate the intensity of a person's emotions.

Tactile Secondary Encodings

Currently, the use of tactile output is primarily limited to Braille devices and highly specialized haptic devices. As a result, many dimensions of tactile information are not regularly used for primary encoding. Therefore, secondary encoding issues have not yet appeared. Many secondary encoding issues may exist and should be considered when tactile encodings become more readily used for primary encoding.

ENCODED INFORMATION

Secondary encodings appear for specific reasons and present different kinds of information. This section will describe three reasons secondary encodings appear, as well as three kinds of information that they convey.

Reasons Secondary Encoding Appear

Computer system and software developers may apply secondary encodings intentionally, incorrectly, or accidentally.

When a person communicates, they often wish to share a certain message. They try to present the information using encodings that are most suitable for their purpose. They attempt to be careful such that the meaning is not misinterpreted. Secondary encodings help enforce the intended meaning and interpretation of the main message. For example, during a video presentation, the presenter deliberately uses eye contact and specific intonations to emphasize certain words.

However, people sometimes use secondary encoding wrongly or insert secondary encodings that contradict their intended meaning. In the presentation example, the presenter may use the wrong intonations, leading to miscommunication. The execution does not have the intended effect or meaning.

Sometimes, a secondary encoding is used but no meaning is intended by it. However, the audience may attach meaning to it. For example, many web designers use color for decoration without any intent to convey meaning. Another example is the presenter could accidentally use hand gestures that offend the audience. Although a person may not intend to offend or cause confusion, it is a common consequence of accidental secondary encodings. This problem occurs only for people within the modality of the encoding.

Information Regarding Emotions

One of the most common uses of secondary encoding is to add emotional information to more objective content. A number of illustrations where secondary encodings are used to present emotional information were discussed above. People often attempt to hide their emotions. However, emotions are typically accidentally inserted into the secondary encodings. The information may naturally appear in the encodings without the person realizing it.

Information Regarding Comfort

Secondary encoding can also provide information on a person's comfort level. In familiar situations, a person will feel more confident, which may be apparent from the way they walk and speak. In unfamiliar situations and environments, the person will feel unsure and less confident. They may be hesitant and be more inclined to speak quietly.

A person's personality will also change based on the how well the two people know each other. During the first encounter, people would act more formally and carefully to provide a good first impression. As they become more familiar, they will become more comfortable and be less formal.

Information Regarding Importance

Secondary encodings can be very effective in assigning importance to specific words in speech and text. It differentiates those words from the rest of the content. As a result,

those words create a greater impact on the audience. Without this information, all of the words will have the same importance and will not be remembered as easily.

Secondary encoding can also be used to explicitly make certain words appear less important, e.g. by using smaller text. This can be misleading. For example, advertisements use small fonts to hide the important details that may ultimately affect whether or not a purchase is to be made. Less emphasized information has less impact and is easily forgotten.

GUIDELINES FOR USING SECONDARY ENCODINGS

The discussions above present problems that need to be overcome before attempting to translate secondary encodings into alternative formats to meet accessibility needs.

This paper would like to propose the following guidelines regarding the use of secondary encoding:
1. If an attribute is used for secondary encoding:
   a. Its only use within a system should be secondary encoding (i.e. it should not also be used accidentally).
   b. It should be used consistently throughout the system (i.e. it should encode the same meaning throughout the system).
2. Information about a set of attributes and their associated meanings that are used for secondary encodings should be available to the user of a system.
3. Where multiple secondary encodings are used, they should not conflict with one another or with the primary encodings.

MAKING SECONDARY ENCODING INFORMATION ACCESSIBLE REVISITED

After identifying the set of secondary encodings and the associated information, the information needs to be inserted into the media file. One possible solution is to use or create XML tags associated with the location, type, and information of the encoding. The tagging system would need to be capable of dealing with dynamic (e.g. video) and static (e.g. picture) media.

Once the tagging system has been established, composition tools are needed for inserting tags into all types of modalities and media. Tools are also needed for extracting the tags from the media file and translating them into the modality that the user can interpret and understand. Development and design of the tagging system and tools are outside the scope of this paper.

However, even with the existence of such technologies, they would be useless if authors of the media files do not use the tools to insert secondarily encoded information. Despite the existence of alt-text for images on a webpage, very few websites use it appropriately, if at all. Without the authors' cooperation, the secondary, and sometimes primary, encoding information may remain missing and inaccessible.

CONCLUSION

The lack of information on how secondary encoding affects accessibility suggests either not many people have knowledge of the problems or people consider the problems to be insignificant. Different types of secondary encodings are generally seen as separate issues when they need to be seen as aspects of the same problem. Developers need to treat secondary encoding as a single problem. By doing so, they would treat all of the problems at the same time instead of focusing on a specific type (such as color) while forgetting the others.

This paper acts as a starting point to centralize information regarding secondary encoding. The issues and solutions presented in this paper apply not only to secondary encoding, but also the primary encoding. Primary encoding is analogous to secondary encoding, therefore, it too has the same problems. When two primary encodings are paired together but only one can be detected, the information from the other encoding is missing. Not having all of the information leads to inaccessibility.

This paper has made two contributions: unifying the concept of secondary encodings and providing guidance on the use of secondary encodings.

REFERENCES

Campbell, N. (2005). Getting to the Heart of the Matter: Speech as the Expression of Affect; Rather than Just Text or Language. *Language Resources and Evaluation, 39*(1), 109-118.

Ekman, P. & Oster, H. (1979). Facial Expressions of Emotion. In *Annual Review of Psychology, 30*(1), 527-554.

Fels, D., Lee, D., Branje, C., Reid, M., Shallow, E., & Hornburg, M. (2005). Emotive Captioning and access to Television. In *Proceedings Americas Conference on Information Systems*. http://www.theclt.org/Papers_New/burnt_toast_omaha.pdf

Festinger, L. (1966). *A Theory of Cognitive Dissonance*. Stanford, California: Stanford University Press.

Harrison, R.P. (1974). *Beyond Words: an introduction to nonverbal communication*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Oudeyer, P. (2003). *The production and recognition of emotions in speech: features and algorithms*. http://www.csl.sony.fr/~py/emotionsIJHCS.pdf

Pennsylvania State University. *Creating Accessible Web Sites*, 2005. http://tlt.its.psu.edu/suggestions/accessibility/font.html

Poyatos, F. (1993). *Paralanguage: a linguistic and interdisciplinary approach to interactive speech and sound*. Amsterdam, Netherlands: John Benjamins Publishing Co.

Sobin, C. & Alpert, M. (1999). Emotion in Speech: the Acoustic Attributes of Fear, Anger, Sadness, and Joy. In *Journal of Psycholinguistic Research, 28*(4), 347-365.

Wicklund, R.A. & Brehm, J.W. (1976). *Perspectives on Cognitive Dissonance*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.