VNU Journal of Foreign Studies

Journal homepage: https://jfs.ulis.vnu.edu.vn/

RESONANCE IN EXPRESSIONS OF FACIAL AFFECT AND VOICE AFFECT IN "FROZEN" ANIMATION

Vo Hoang Chau*

Department of Foreign Languages, Quy Nhon University,
170 An Duong Vuong, Quy Nhon City, Binh Dinh Province, Vietnam

Received 22 August 2024
Revised 25 March 2025; Accepted 13 May 2025

Abstract: In contemporary discourse analysis, understanding communication extends beyond written and spoken words to encompass a broader range of modes, including images, gestures, and sounds. Multimodal Discourse Analysis emerges to explore how meaning is constructed and communicated through these diverse modes within various contexts. This study addresses this fruitful field by examining the relationship between voice qualities and facial expressions in conveying emotions through a popular children's animation called "Frozen". The analysis is based on Ngo et al.'s (2022) framework to explore the interplay of the two systems of Facial Affect and Voice Affect from an interpersonal perspective. In analyzing Elsa, one of the two main characters in the animation, we found that she is largely portrayed as having low spirits, evident in the type of facial and vocal features she expresses. Although there are occasional disjunctions between these two systems of affect, likely stemming from contextual variables or production flaws, the majority of instances reveal resonance between Elsa's expressed features of Facial Affect and Voice Affect. Theoretically, the findings involving a complete animation contribute to the literature on multimodal texts targeted at children. It also yields practical benefits in encouraging the utilization of multimodal resources to enhance children's language and social development.

Keywords: systemic functional semiotics, multimodal discourse analysis, children's animations, Facial Affect, Voice Affect

* Corresponding author.
Email address: vohoangchau@qnu.edu.vn
https://doi.org/10.63023/2525-2445/jfs.ulis.5390

SỰ CỘNG HƯỞNG GIỮA NÉT MẶT VÀ GIỌNG NÓI ĐỂ THỂ HIỆN CẢM XÚC TRONG PHIM HOẠT HÌNH "FROZEN"

Võ Hoàng Châu

Khoa Ngoại ngữ, Trường Đại học Quy Nhơn,
170 An Dương Vương, Thành phố Quy Nhơn, Tỉnh Bình Định, Việt Nam

Nhận bài ngày 22 tháng 8 năm 2024
Chỉnh sửa ngày 25 tháng 3 năm 2025; Chấp nhận đăng ngày 13 tháng 5 năm 2025

Tóm tắt: Trong phân tích diễn ngôn hiện đại, bên cạnh chữ viết và lời nói, việc giao tiếp được thực hiện đa dạng hơn qua các kênh như: hình ảnh, cử chỉ và âm thanh. Phân tích diễn ngôn đa phương thức ra đời nhằm tìm hiểu cách ý nghĩa được hình thành và truyền đạt qua những kênh khác nhau trong những ngữ cảnh khác nhau. Nghiên cứu này được tiến hành để đánh giá mối quan hệ giữa giọng nói và khuôn mặt trong việc biểu đạt cảm xúc qua phim "Frozen", một bộ phim hoạt hình nổi tiếng dành cho trẻ em. Sự giao thoa giữa các yếu tố biểu cảm khuôn mặt và biểu cảm giọng nói được phân tích theo góc nhìn liên nhân dựa trên khung lý thuyết của Ngô và các cộng sự (2022). Khi phân tích Elsa, một trong hai nhân vật chính của bộ phim, chúng tôi phát hiện rằng nhân vật này phần lớn thời gian được thể hiện là một người u sầu qua những đặc điểm về khuôn mặt và giọng nói. Tuy có sự chênh lệch trong một vài trường hợp do ngữ cảnh hoặc lỗi kỹ thuật trong quá trình sản xuất phim, đa phần các biểu cảm khuôn mặt và biểu cảm giọng nói của Elsa có sự cộng hưởng để truyền đạt cảm xúc nhân vật. Về mặt lý thuyết, bằng việc phân tích một bộ phim hoàn chỉnh, các kết quả này góp phần làm phong phú thêm nghiên cứu về văn bản đa phương thức dành cho trẻ em. Bên cạnh đó, bài viết khuyến khích ứng dụng các nguồn tài nguyên đa phương thức nhằm tăng cường sự phát triển ngôn ngữ lẫn nhận thức xã hội của trẻ nhỏ.

Từ khóa: ký hiệu học chức năng hệ thống, phân tích diễn ngôn đa phương thức, phim hoạt hình trẻ em, biểu cảm khuôn mặt, biểu cảm giọng nói

1. Introduction

In the dynamic landscape of communication, meaning is seldom confined to mere words. In our current society, communication transcends traditional linguistic boundaries, intertwining with a plethora of visual, auditory, spatial, and gestural elements. It is within this tapestry of modes that Multimodal Discourse finds its relevance and significance. The groundwork for multimodal research was laid out by O'Toole (1994) and Kress & van Leeuwen (2006), drawing upon Halliday's (1985) social semiotic approach to language, which conceptualizes words, sounds, and images as interconnected systems and structures with inherent meaning potential (O'Halloran, 2004). Studies involving multimodal discourse analysis draw their research subjects from a variety of sources, such as textbooks (Unsworth & Ngo, 2015; Silva, 2016), picture books for children (Vo, 2016; Ton, 2019), advertisements (Akmal et al., 2022; Ruswardiningsih & Djohan, 2022), music videos (Brady, 2015; Sánchez-Vizcaíno et al., 2023), political cartoons (Mowafy, 2021; Wu, 2023) and movies (Lisiecka, 2019; Nurlina, 2023). These sources are referred to as "multimodal phenomena", in which various "semiotic resources", for example, language, images, music and gestures, are used to create and clarify meaning (O'Halloran, 2004, pp. 121-122).

This study focuses on a popular multimodal phenomenon, movies, or specifically

animated movies for children, from a Systemic Functional Linguistics (SFL) perspective. Research on children's animations from an SFL perspective primarily follows two directions. The first approach is utilizing Systemic Functional Grammar (SFG) (Halliday, 1994) and the Grammar of Visual Design by Kress & van Leeuwen (2006) to describe the characteristics of still shots taken from animations with regards to the three metafunctions (Phan, 2020; Rady, 2023). The second approach also applies the two above-mentioned theories, but the core purpose is to discover how language and its corresponding visual element are combined to portray specific ideological codes, values and stereotypes (Asseel, 2020; Rasheed et al., 2020; Shehatta, 2020). Most notably, the studies only concentrate on static images, whereas the semiotic resources in animations integrate across both visual and auditory modalities.

Our study aims to bridge this gap by conducting an in-depth analysis of the relationship between sounds and accompanying facial expressions of the main character in a famous children's animation, "Frozen", when conveying emotions. The framework used in this study was designed by Ngo et al. (2022) for analyzing paralanguage using Systemic Functional Semiotics. Research questions for this study are:

1. What types of Facial Affect and Voice Affect are used in the visual and spoken discourse featuring the main character Elsa in "Frozen" animation?

2. How do the expressions of Facial Affect and Voice Affect resonate one another in "Frozen"?

This choice is driven by both theoretical and practical considerations. Theoretically, the analysis seeks to offer empirical insights into this relatively novel field of research. Traditionally, facial expressions have been given significantly more attention than vocal ones when it comes to portraying emotions (Scherer, 1986, cited in Juslin & Laukka, 2003; Scherer & Banse, 1991), although "with many kinds of animals, man included, the vocal organs are efficient in the highest degree as a means of expression" (Darwin, 1897, p. 88). However, as Sauter (2006) pointed out, one drawback of using speech to identify emotion is that it contains semantic information; this piece of information might be "congruent or incongruent with the paralinguistic content of the signal, and hence interfere or facilitate the listener's judgment of the emotional tone of the speech" (p. 28). Therefore, instead of analyzing speech or facial expressions separately, we decided to employ both paralinguistic means into our study to better interpret the research subject's emotions. To the best of our knowledge, there has been no prior research exploring identical or similar data in this direction. From a practical standpoint, the significance of the analysis lies in acknowledging the value of animations as sources of entertainment and as useful educational tools that can foster emotional, social and language development in children.

This paper is structured as follows: First, the theoretical framework for analyzing paralanguage is presented. This is followed by a detailed description of the methodologies employed to conduct the study. Subsequently, the major findings are presented and discussed. The paper is concluded with the significance of the research in multimodal discourse analysis as well as implications for further studies.

2. Literature Review

2.1. Systemic Functional Linguistics

Systemic Functional Linguistics, first introduced by Halliday in the early 1960s, is a theory that views language as a social semiotic resource (Bateman, 2008). "The value of a theory lies in the use that can be made of it, and I have always considered a theory of language

to be essentially consumer oriented" (Halliday, 1985, p. 7). Language transcends arbitrary rules and structures; it is a tool people use to create meaning and interact with the world. Within SFL, there is the theory of SFG, in which Halliday (1994) introduced three modes of meaning that operate simultaneously in all utterances; they are the ideational (experiential and logical), interpersonal and textual metafunctions. The ideational metafunction refers to the way language is used to convey information about the world and represent experiences. The interpersonal metafunction is utilized to encode social relationships and express our attitudes. Finally, the textual metafunction uses language to create coherent and cohesive texts (Butt et al., 2003). Derived from SFG, Kress and van Leeuwen (2006) proposed a systemic and comprehensive framework for visual analysis in "Reading Images: The Grammar of Visual Design". The grammar of images also has three layers of meanings that correspond to the three metafunctions: representational meaning, interpersonal meaning and compositional meaning.

According to Ngo et al. (2022), an SFL-based approach to studying paralanguage proves advantageous for various reasons, three of which are directly relevant to our study. Firstly, the expansion of SFL-inspired research that extends beyond language to encompass various modalities, such as behaviour, sounds and music, underscores the need for a more extensive and unified theoretical framework that is not restricted to only language and images. Secondly, paralanguage and spoken language are so closely intertwined and consistently involved in various intermodal texts that the necessity of a shared metalanguage becomes evident. Lastly, work on Appraisal (Martin & White, 2005), the language of evaluation in SFL, is inextricably linked to systems used to portray emotions though paralinguistic features.

2.2. Paralanguage

In the given framework, paralanguage refers to "semiosis dependent on language and realized through both voice quality and body language (including facial expression, gesture, posture and body movement)" (Ngo et al., 2022, p. 3). Unlike somasis, also known as non-semiotic behavior, semiosis concerns systems of signs, or systems of behavior with embodied meaning.

Zappavigna and Martin (2018) devised a model for dealing with two dimensions of the relation between language and paralanguage; these dimensions are "linguistic body language" and "epilinguistic body language". The terminologies were later revised in Ngo et al.'s framework and changed to "sonovergent paralanguage" and "semovergent paralanguage" respectively (2022, p. 22). Sonovergent paralanguage is the kind of paralanguage that converges with prosodic phonology, that is, rhythmic and intonation patterns of spoken language (Nespor & Vogel, 2007). Regarding semovergent paralanguage, the focus of our study, it is the type that converges with lexico-grammar and semantics (the meaning-making resources) of spoken discourse (Martin & Rose, 2007).

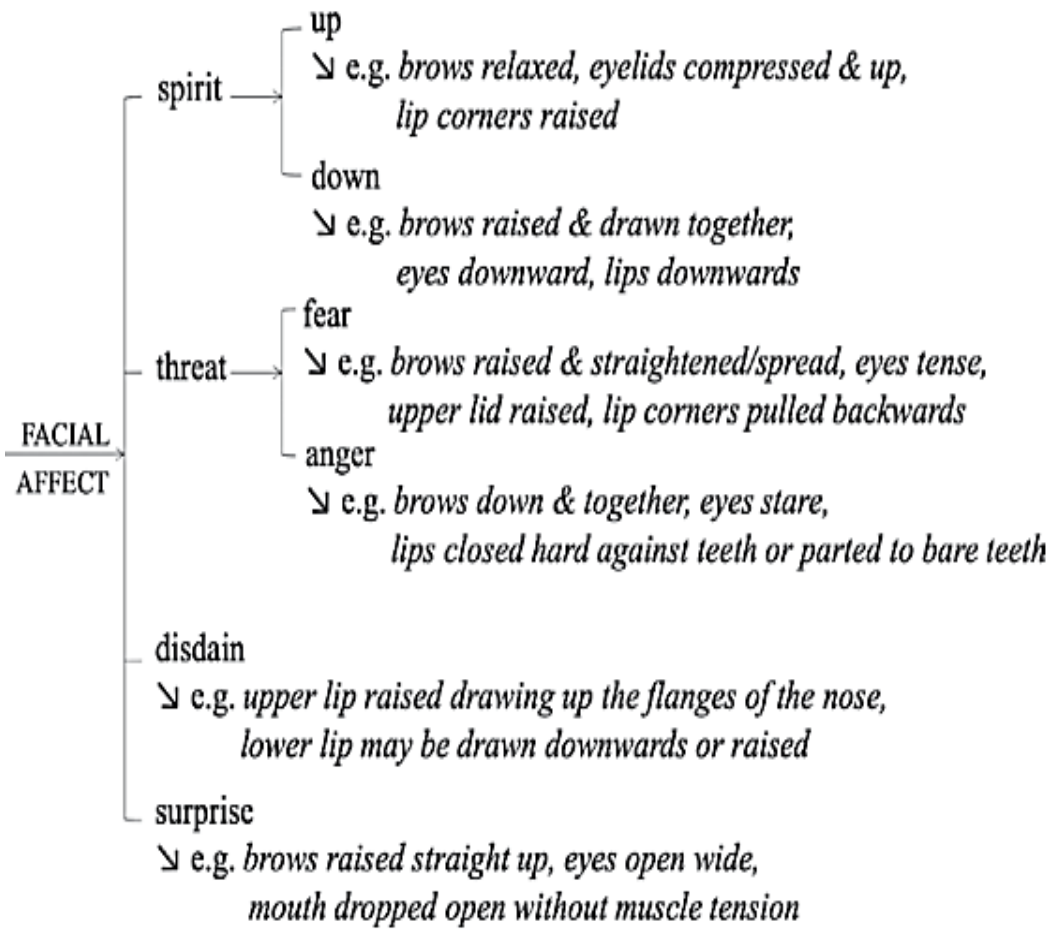
With an aim of analyzing the emotions of a protagonist in the "Frozen" animation, this study centers on semovergent paralanguage inspected from the interpersonal perspective as interpersonal semovergent paralanguage potentially resonates with Appraisal resources, specifically the sub-system of Affect used for communicating emotions in the system of Attitude. Paralinguistic Affect portrays expressions of emotion in Facial Affect and Voice Affect (Ngo et al., 2022).

Features of the system of Facial Affect are realized through muscle movements of the face, namely eyebrows, eyelids, eyes, mouth and lips. This system takes into account the contributions made by previous researchers, Darwin (1872), and Ekman and colleagues (Ekman

& Friesen, 2003; Ekman, 2004, cited in Ngo et al., 2022). Darwin created the principle of antithesis, in which he proposed that opposing facial movements lead to opposing emotions. Ekman introduced the six basic universal facial expressions of emotion, which are happiness, sadness, anger, fear, disgust and surprise.

Figure 1

The System of Facial Affect (Ngo et al., 2022, p. 121)



The network in Figure 1 models Facial Affect with six features of emotions: [spirit:up], [spirit:down], [fear], [anger], [disdain] and [surprise]. Each feature is shown through a different set of facial movements. [spirit:up] and [spirit:down] are opposing features of [spirit], demonstrating happiness and sadness; Ngo et al. (2022) did not name these features "happiness" and "sadness" to avoid confusion with the feature of [happiness] in the linguistic system of Attitude.

While language allows for nuanced expressions of emotion through lexical variations, such as "happy", "excited" "pleased" and "ecstatic", paralinguistic expressions have a relatively limited range for expressing emotions. Therefore, in analyzing intermodal resonance, while linguistic expressions may convey subtle differences, facial expressions remain more generalized, and one given expression of Facial Affect may have to couple with a wide range of lexical realizations.

Figure 2

The System of Voice Quality (Ngo et al., 2022, p. 130)

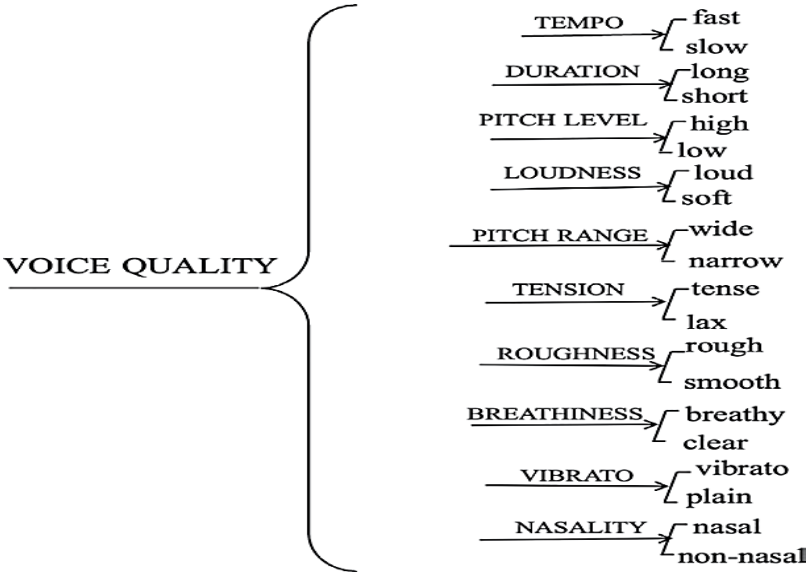
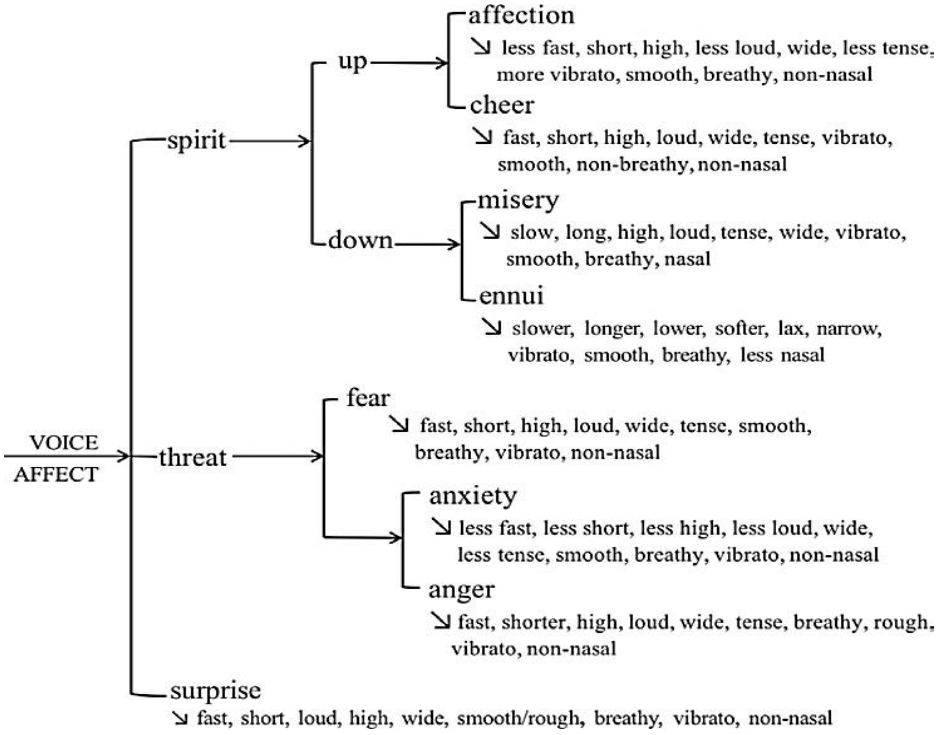


Figure 3

The System of Voice Affect (Ngo et al., 2022, p. 131)



Another semiotic resource utilized for conveying emotions is the voice. Additional examination of voice qualities enables us to make finer comparisons in paralinguistic emotions than solely relying on facial expressions. The system of Voice Quality in the framework was developed based on van Leeuwen's "sound quality systems" (1999, p. 151). Van Leeuwen's model includes subsystems of Pitch level, Pitch range, Loudness, Tension, Roughness,

Breathiness, Vibrato and Nasality. The Pitch level can be high or low, depending on whether speakers use the higher or lower end of their pitch range. The Pitch range, which can be wide or narrow, is the difference between the highest pitch and the bottom level that the pitch does not go below. Next, the Loudness of the voice indicates how much energy is put into it and is measured in dB. With Tension, when a person's voice becomes tense, it becomes higher, sharper, and particularly, more tense due to tense throat muscles. When it comes to Roughness, a rough voice is one with hoarseness, harshness and rasp, while a smooth voice is clean, smooth, and "well-oiled" (p. 132). Moreover, when a voice is breathy, it becomes soft, and the sounds of breath mix in the tone of voice itself. Vibrato indicates the vibration of the voice; the voice could exhibit some kind of irregular wavering or trembling or it could sound plain and unmoved. The last subsystem is Nasality. It is defined as "sound produced with the soft palate lowered and mouth unblocked, so the air escapes both via the nose and via the mouth" (p. 135); thus, the voice appears blocked and tense. To the van Leeuwen's system, Ngo et al. (2022) added the subsystems of Duration with time-related features of [long] and [short], and Tempo with [fast] and [slow]. However, in this study, we have omitted Duration because Ngo et al. did not provide a detailed explanation of this subsystem or a reliable way to determine whether an utterance is [long] or [short].

As van Leeuwen noted, "Every sound quality is a mixture of different features. The voice is never only high or low, or only loud or soft, or only tense or lax" (p. 131). This means that in a single utterance, several voice qualities occur simultaneously, and each feature in the Voice Affect system in Figure 3 is a particular mixture of several qualities.

3. Methodology

3.1. Research Method

To address the research questions, we employed both the qualitative and quantitative methods. The qualitative method was first used to identify the types of Facial Affect and Voice Affect present in the spoken and visual discourse of "Frozen"; this was performed through direct observation (Facial Affect), and perceptual examination and the use of Pratt, a speech analysis software (Voice Affect). Consequently, the data was inputted into Microsoft Excel to be quantitatively analyzed to arrive at the occurrence frequencies of different features. Finally, the qualitative method was utilized to investigate the resonance between these two systems of Paralinguistic Affect. To do this, we compared the types of Voice Affect and Facial Affect that each tone group exhibits to see whether they belong to the same branch of emotions.

3.2. Data Description

The animation that serves as the data for this study is "Frozen", downloaded from fsharetv.com. "Frozen" is a Disney animated movie released in 2013. The story revolves around two princesses, Elsa and Anna, who live in the kingdom of Arendelle. Elsa has magical powers but she struggles to control them. To protect her sister and the kingdom, Elsa isolates herself, causing a rift between the sisters. Upon Elsa's coronation as queen, her powers are revealed, and she inadvertently plunges Arendelle into eternal winter. Feeling responsible, Elsa flees to the mountains, leaving her sister to search for her. Throughout the journey, Elsa grapples with her powers and her duty as queen, ultimately embracing her true self and learning the power of love and acceptance. Of the many animations for children, we decided to choose "Frozen" because of its overwhelming popularity, illustrated by its massive commercial success; the animation grossed $1.28 billion USD worldwide (USA Today, 2023). It also won two Academy

Awards for Best Animated Feature and Best Original Song (Playbill, 2014). The animation has become a beloved classic in the Disney canon, inspiring merchandise and igniting widespread fan enthusiasm.

Out of the two main characters in "Frozen", Elsa and Anna, we decided to focus on Elsa. As a character who experiences a wide range of emotions throughout the movie, Elsa offers ample opportunities for studying the correlation between vocal intonation and emotional states. In addition, Elsa undergoes significant character development as the story unfolds, which may be reflected in changes in her voice quality and facial expressions. Another reason is that we wanted to study a character in all of the movie, instead of selecting only the prominent scenes. Due to the time constraint and the scope of this study, it is unrealistic to analyze Anna as she appears in almost every scene, whereas Elsa's screen time is neither limited nor abundant.

3.3. Data Analysis

In this study, the unit of analysis for the visual discourse is single static images extracted from the movie. The still capture taken from a particular scene is the one that shows the most intense amount of emotion during that scene. The muscle movements of the face are compared with the descriptions in the system of Facial Affect to identify which feature the image portrays.

The unit of analysis for the spoken discourse is the tone group. The tone group is the smaller component of the utterance; an utterance may consist of one or more tone groups. The boundaries of tone groups are often marked by pauses in speech – that is, a tone group ends and is followed by a pause, and another tone group begins following the pause (Roach, 2009).

Analyzing the spoken discourse was relatively more difficult and time-consuming than analyzing the visual one. First, the neutral speech of Elsa was taken from a scene in which she expresses the mildest of emotions. This provided a baseline of voice qualities against which the variations of each feature in Figure 3 can be compared to define what feature a tone group belongs to. Next, the audio extractions underwent perceptual examination of voice qualities by the researchers. The results of this process were then compared with Pratt visualizations to reach the final conclusions. Pratt is an open-source tool created by Boersma and Weenink to analyze speech in phonetics, and it has been developing since 1992 (Boersma & van Heuven, 2001). However, as Ngo et al. (2022) mentioned, using technology to measure the acoustic qualities of voice may not be entirely accurate. This is because in movies, the characters' voices are often recorded with background noise and music, and in such cases, the visualizations might calculate and present values of all the sounds in one measurement. Therefore, the perceptual analysis is considered to be the more reliable method in this study.

4. Findings and Discussion

4.1. Types of Facial Affect and Voice Affect in "Frozen"

A total of 156 tone groups (in 111 utterances) and 156 static images were collected from Elsa's scenes in "Frozen". Each tone group/image was coded with a number from (1) to (156) within their system of Facial Affect and Voice Affect, making it easier for the researchers to present the findings.

Regarding the system of Facial Affect, the character Elsa portrays five out of six features; no instance of [disdain] was recorded throughout the movie. Next, there are four different situations in which the images are classified as "Others". Firstly, the facial expression of the character cannot be identified because her face is largely or entirely covered by another character (e.g. (14) and (16)), she has her back turned to the viewer (e.g. (49) and (73)), or her

face is captured at a long-shot angle (e.g. (80)). Secondly, Elsa expresses a mixture of emotions, for example, a mixture of [spirit:up] and [anger] (image (3)) or a mixture of [spirit:down] and [fear] (image (13)). Thirdly, the muscle movements are not significant enough to categorize the image into a specific feature, for instance, in image (20), Elsa's eyebrows are only slightly drawn together and the mouth corners are slightly pulled downwards, thus, the signs are not enough for it to express [spirit:down]. Lastly, if her emotion is neutral, or only one out of the three important parts of the face (eyebrows, eyes and mouth) shows movements (e.g. (33) and (83)), the image is placed into the "Others" group.

Figure 4

Example of an Image Belonging to "Others"

Table 1

Types of Facial Affect in "Frozen"

No.	Types of Voice Affect	Number of Occurrences	Percentage
1	[spirit:up]	29	18.6%
2	[spirit:down]	61	39.1%
3	[fear]	11	7.1%
4	[anger]	6	3.8%
7	[disdain]	0	0%
8	[surprise]	3	1.9%
9	Others	46	29.5%
	Total	156	100%

Table 1 demonstrates that among the five features, Elsa most frequently expresses the [spirit:down] emotion with her eyebrows raised and drawn together, eyes cast downwards, lower eyelids slightly raised, and mouth corners pulled downwards. However, it is found that the [spirit:down] feature can also be expressed with slightly different movements from the characteristics described in Figure 1. Instead of casting her eyes downwards, Elsa shows even stronger emotions by closing her eyes (e.g. (63), (90) and (112)). Elsa's eyebrows may also be drawn down instead of raised up; this is observed in more than 30 instances (e.g. (54) and (118)).

Figure 5

Elsa Expresses [Spirit:Down]

The second most common feature is [spirit:up], but the number of occurrences is only half that of [spirit:down]. This means that, during the majority of her screen time, Elsa is depicted as a young person who lives in sadness and misery. An interesting point to consider is that when expressing [spirit:up], Elsa's eyebrows are often not relaxed like the description, but instead, are raised (e.g. (151) and (156)).

Figure 6

Elsa's Expression of [Spirit:Up]

Although there are some differences, the researchers still chose to put the aforementioned cases into the [spirit:down] or [spirit:up] category based on the contextual clues of the scenes. As Ngo et al. (2022) remarked, when analyzing from a social semiotic perspective, one should "always keep meaning in mind" (p. 76).

Turning to Voice Affect, it is worth highlighting that while each feature consists of options belonging to the nine sub-systems of Voice Quality, in most of the audio files collected as data, the characteristics of the voice do not correspond with all nine sub-systems. However, the voice of a tone group is still assigned a particular feature if its qualities match the options of at least six out of the nine sub-systems in the description.

Table 2

Types of Voice Affect in "Frozen"

No.	Types of Voice Affect	Number of Occurrences	Percentage
1	[affection]	13	8.3%
2	[cheer]	9	5.8%
3	[misery]	24	15.4%
4	[ennui]	2	1.3%

5	[fear]	14	9.0%
6	[anxiety]	30	19.2%
7	[anger]	6	3.8%
8	[surprise]	5	3.2%
9	Others	53	34.0%
	Total	156	100%

Table 2 presents the results of the Voice Affect analysis. Similar to Facial Affect, there are certain cases that do not match with the eight given features and thus, need to be grouped together in "Others". The voice qualities of these cases either match that of Elsa's neutral speech (e.g. (27) and (37)) or only some aspects (less than six) of voice qualities align with the descriptive terminologies (e.g. (19) and (32)). For example, in order for a tone group to express [misery], it needs to be slow, high, loud tense, wide, vibrato, smooth, breathy and nasal. The voice in audio (19) is slow, nasal, smooth, but it is not tense, high, wide, or is vibrato enough to be classified as [misery]. Among the eight features, the most prevalent is [anxiety], followed by [misery]. This result further reinforces the conclusion that Elsa's mood is indeed low for most of the time.

4.2. Resonance in Expressions of Facial Affect and Voice Affect in "Frozen"

In the previous section, we have individually analyzed two equally important paralinguistic systems pertaining to the conveyance of emotion: Facial Affect and Voice Affect. While animated facial expressions vary in the movements of only three major expressive areas – namely, eyes, eyebrows, and mouth – the voice exhibits diversity across a broader spectrum of vocal qualities, resulting in a wider variety of potential groupings. Therefore, when investigating emotions, it is necessary to place the two systems side by side to enhance the accuracy of analysis. "If we accept that natural (i.e. not performed) vocal and facial expressions of emotion are biological in nature" (Darwin, 1872, cited in Ngo et al., 2022), it is reasonable to expect resonance across the systems of Facial Affect and Voice Affect in all cases (except in instances of deliberate voice manipulation, such as in showing sarcasm).

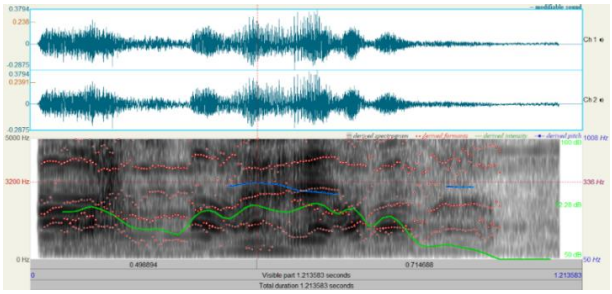
With regards to the resonance in expressions of Facial Affect and Voice Affect in "Frozen", the results can be divided into three groups. In the first group, options in Facial Affect and Voice Affect are found to resonate with each other. The number of instances recorded for this group is 58. We can take tone group (5) as an example. In a convergent speech, Elsa tells Anna to watch closely as she uses her superpowers to conjure up snowflakes filling the entire ballroom. Her Facial Affect expresses [spirit:up] as realized through raised eyebrows, slightly compressed eyelids due to raised cheeks and raised lip corners forming an open-mouth smile (Figure 7).

Figure 7

Elsa Performing Magic

Figure 8

Pratt Visualization of Tone Group (5)



Her voice also realizes [cheer] as it is vibrato, high and loud; it has a maximum pitch level of 336 Hz and intensity of 72 dB, much higher than that of her neutral utterances, which are 202 Hz and 59 dB. It is also relatively [smooth] but [tense] (illustrated respectively in the Pratt visualization by the absence of sub-harmonics in the sound waveform and the darkness of the upper bands in the grey area). The spectrogram does not show any evidence of breathiness as the darkness remains consistent in both upper and lower areas (Figure 8).

As for the second group, options in Facial Affect and Voice Affect do not resonate with each other. The number of cases recorded for this group is 25. Tone group (98) can be seen as a typical example. In this scene, Elsa and Anna are arguing inside the ice castle, then suddenly, Olaf, the snowman that the two sisters built when they were little, comes in. Elsa's voice at this moment is less loud than high-level [cheer] (59 dB), lax, high (maximum pitch of 370 Hz), smooth, breathy (shown in the Pratt visualization with a grey area that has a lighter upper part and darker lower part) and vibrato; thus, this tone group expresses [affection] (Figure 9).

Figure 9

Pratt Visualization of Tone Group (98)

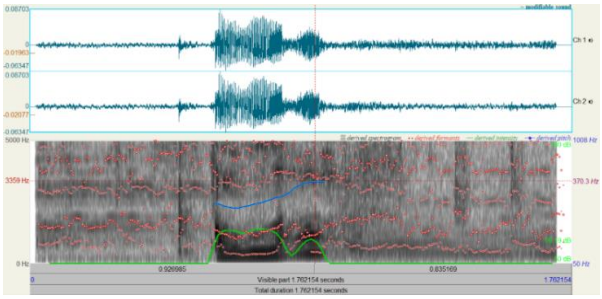


Figure 10

Elsa seeing Olaf

We might expect her facial expression to convey [spirit:up] when she sees the creature that she created and loved. However, in this instance, her eyebrows are drawn down and together, eyes slightly cast downwards, and corners of mouth pulled downwards, which are characteristics of [spirit:down] (Figure 10).

However, this dissonance would seem natural if we consider the scene preceding it; Elsa and Anna have been caught up in an argument. Consequently, when Elsa catches sight of Olaf, her face still maintains its [spirit:down] state (evident in image (96)) for some time before transitioning to [spirit:up] (evident in image (98)). We should also bear in mind the fact that animations present a simulated reality with imperfections in character depiction, the presence of background music and other ambient sounds as well as differences in microphone proximity for voice actors. Therefore, where such disjunction between Facial Affect and Voice Affect cannot be interpreted as meaningful, it may simply be explained as a slip in the production process (Ngo et al., 2022).

The final group comprises Facial and Voice combinations that come between having resonance and exhibiting disjunction. The paralinguistic features in these instances (a total of 73 instances) mainly belong to the "Others" group.

4. Conclusion

This study examines the types of Facial Affect and Voice Affect expressed by the character Elsa and how these paralinguistic expressions resonate with each other. In analyzing Elsa's expressions in "Frozen", it is evident that she predominantly conveys a sense of low spirits, characterized by specific facial and vocal cues. While her expressions often align with predefined categories, variations do exist. Despite some discrepancies between facial and vocal expressions, the majority of instances demonstrate resonance between Facial Affect and Voice Affect systems.

Considering the ongoing advancements in multimedia and the expanding influence of English as a global language, the significance of the findings becomes increasingly apparent. An understanding of the various meaning-making resources and their interconnections through animations will undoubtedly be advantageous for children. Animations provide an immersive language learning experience, allowing children to hear authentic English spoken in various contexts, accents, and tones. At the same time, there are visual cues that aid in understanding language and context, making it easier for children to grasp new vocabulary and expressions. Additionally, movies often depict cultural aspects and societal norms, providing valuable cultural insights alongside language learning. This study has also helped to expand the description of the Facial Affect [spirit:down] and [spirit:up] features as well as providing reasons explaining for the disjunction of speech and facial expressions in certain cases. These details may be useful for further research along the same lines.

Despite its contributions, the study has obvious limitations. The most apparent limitation is the size of the data, which involves solely one animation. We hope that future studies with the same framework will be conducted with larger datasets to ascertain whether the findings presented in this study can be generalized to encompass the entire genre under investigation. Another limitation lies in the fact that we only analyzed the relationship between the voice and its accompanying facial expressions, whereas, interpersonal semovergent paralanguage also involves posture, muscle tension, the position and movements of the hand/arm (Hood, 2011). Future researchers could consider investigating these paralinguistic elements and their interplay. Another suggestion is analyzing paralanguage from other

perspectives, namely the ideational and textual perspective. Finally, as we are currently surrounded by great varieties of multimodal texts, it is beneficial to further investigate more practical implications of multimodal discourse with the aim of maximizing the use of these resources in fostering language education and social awareness.

References

Akmal, A., Latiffani, C., & Almeina, I. (2022). An analysis of multimodals on Grab advertisement. *Jurnal Ilmu Pendidikan Nonformal, 8*(1), 65-74. https://doi.org/10.37905/aksara.8.1.65-74.2022

Asseel, D. A. (2020). *Seeing the unseen: Euphemism in animated films a multimodal and critical discourse study* [Unpublished M.A. Thesis]. Lancaster University.

Bateman, J. (2008). *Multimodality and genre: A foundation for the systematic analysis of multimodal documents*. Springer, Hampshire.

Boersma, P., & van Heuven, V. (2001). Speak and unspeak with Praat. *Glot International, 5*(9), 341-347.

Brady, C. K. (2015). *A multimodal discourse analysis of female K-pop music videos* [Unpublished Ph.D. Thesis]. University of Birmingham.

Butt, D., Fahey, R., Feez, S., Spinks, S., & Yallop, C. (2003). *Using functional grammar: An explorer's guide* (2nd ed.). NSW: Macquarie University.

Darwin, C. (1872). *The expression of emotion in man and animals*. London: Harper Collins Publishers.

Gioia, M. (2014, March 19). *Disney's Oscar-winning film "Frozen" will become highest-grossing animated film.* Playbill. https://playbill.com/article/disneys-oscar-winning-film-frozen-will-become-highest-grossing-animated-film-com-216182

Halliday, M. A. K. (1985). Systemic background. In J. D. Benson, & W. S. Greaves (Eds.), *Systemic perspectives on discourse: Selected theoretical papers from the 9th international systemic workshop*. Ablex Publishing Corporation.

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). Edward Arnold.

Hood, S (2011). Body language in face-to-face teaching: A focus on textual and interpersonal meaning. In S. Dreyfus, S. Hood & M. Stenglin (Eds.), *Semiotic margins: Meaning in multimodalities*. Continuum.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code?. *Psychological Bulletin, 129*(5), 770-814. https://doi.org/10.1037/0033-2909.129.5.770

Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). London: Routledge.

Lisiecka, A. (2019). Comparing multimodal film texts: The case of the movie Fame (1980) and its remake Fame (2009). *CROSSROADS A Journal of English Studies, 25*(2), 52-65. https://doi.org/10.15290/cr.2019.25.2.03

Makram, N. M. (2021). A multimodal discourse analysis of American animated feature film (The Wizard of Oz) and Japanese anime (Spirited Away). *Journal of Humanities, Social Sciences and Education, 8*(3), 59-81. https://doi.org/10.21608/buhuth.2021.94743.1186

Martin, J. R., & Rose, D. (2007). *Working with discourse: Meaning beyond the clause*. Continuum, London.

Martin, J. R., & White, P (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan, Basingstoke.

Mowafy, M. M. (2021). US foreign policy during the presidency of Trump: A multimodal discourse analysis of selected political cartoons in Al-Ahram weekly. *BSU International Journal of Humanities and Social Science, 4*(1), 172-203. https://doi.org/10.21608/buijhs.2022.225722

Mulroy, C. (2023, December 20). Highest grossing animated movie of all time: Why it's a debate between two Disney hits. *USA Today*. https://www.usatoday.com/story/life/2023/05/12/highest-grossing-animated films/11729384002/

Nespor, M., & Vogel, I. (2007). *Prosodic phonology: With a new foreword*. Walter de Gruyter, Berlin.

Ngo, T., Hood, S., Martin, J. R., Painter, C., Smith, B. A., & Zappavigna, M. (2022). *Modelling paralanguage using systemic function semiotics.* Bloomsbury Publishing Plc.

Nurlina, N. (2023). *A multimodal discourse analysis on the representation of racism in the film "Armageddon Time"* [Unpublished M.A. Thesis]. Lancester University.

O'Halloran, K. L. (2004). *Multimodal discourse analysis: Systemic-Functional perspectives*. Continuum, London.

O'Toole, M. (1994). *The language of displaced art*. Leicester University Press.

Phan, T. K. H. (2020). *A study of cartoon trailers in English from the perspective of multimodal discourse analysis* [Unpublished M.A. Thesis]. Da Nang University.

Rady, R. M. A. (2023). A multimodal discourse analysis of the American animated feature Film (Pocahontas) by Walt Disney Production. *Egyptian Journals, 6*(44), 1-29. https://doi.org/10.21608/jfhsc.2023.302240

Rasheed, S., Fayyaz, A., & Danish, M. H. (2020). Acquirements of moralistic values through animation: A multimodal analysis of short film "Spell Bound". *European Academic Research, 6*(6), 2917-2937.

Roach, P. (2009) *English phonetics and phonology: A practical course* (4th ed.). Cambridge University Press.

Ruswardiningsih, D., & Djohan, R. S. (2022). An Aqua advertisement's multimodal discourse analysis. *International Journal of Linguistics, Literature and Translation, 5*(1), 230-237. https://doi.org/10.32996/ijllt.2022.5.1.27

Sánchez-Vizcaíno, M., Cancelas-Ouviña, L., & Fonseca-Mora, M. C. (2023). A multimodal analysis for an updated vision of women in music videos. *Observatorio Journal, 17*(3), 171-193. https://doi.org/10.15847/obsobs17320232229

Sauter, D. (2006). *An investigation into vocal expressions of emotions: The roles of valence, culture, and acoustic factors* [Unpublished Ph.D. Thesis]. University College London.

Scherer, K. R., & Banse, R. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion, 15*(2), 123-148. https://doi.org/10.1007/bf00995674

Shehatta, S. (2020). Breaking stereotypes: A multimodal analysis of the representation of the female lead in the animation movie Brave. *Journal of English and Comparative Studies, 2*(1), 170-194. https://doi.org/10.21608/ttaip.2020.133423

Silva, M. d. (2016). Accounting for multimodality in an EFL textbook: Analyzing activities and suggesting ways to approach multimodal texts. *Diálogo das Letras, Pau dos Ferros*, *5*(2), 92-109.

Ton, N. M. N., & Nguyen, T. M. P. (2019). Exploring text-image relations in English comics for children: The case of "Little Red Riding Hood". *VNU Journal of Foreign Studies, 35*(3), 128-138. https://doi.org/10.25073/2525-2445/vnufs.4372

Unsworth, L., & Ngo, T. T. B. (2015). The role of images in Vietnamese textbooks for the teaching of English as a foreign language. *Language and Life, 1*(231), 93-99.

Van Leeuwen, T. (1999). *Speech, music, sound*. Macmillan, London.

Vo, T. M. T. (2016). *English nursery rhymes from a multimodal discourse analysis perspective* [Unpublished M.A. Thesis]. Quy Nhon University.

Wu, A. (2023). The images of China in political cartoons of foreign media. *Open Journal of Modern Linguistics, 13*, 523-558. https://doi.org/10.4236/ojml.2023.134033

Zappavigna, M., & Martin, J. R. (2018). *Discourse and diversionary justice: An analysis of Youth Justice Conferencing*. Palgrave Macmillan, London.

Appendix