

Cognition, Communication, Discourse, 2026, 32, 133-144  
<http://sites.google.com/site/cognitiondiscourse/home>

<https://doi.org/10.26565/2218-2926-2026-32-09>

Received 15.04.2026; revised May 20, 2026;  
accepted May 25, 2026; published online May 31, 2026

## SOUND IN THE MULTIMODAL CONSTRUCTION OF HEALTH AND CONVENIENCE: A CASE STUDY OF SWEETGREEN'S INSTAGRAM ADVERTISING

**Andrii Rybalchenko**

Assistant professor,

Taras Shevchenko National University of Kyiv

(60 Volodymyrska, Kyiv, 01033, Ukraine);

e-mail: [andrei.uni@live.com](mailto:andrei.uni@live.com)

ORCID: <https://orcid.org/0009-0007-1595-7429>

**Article citation:** Rybalchenko, A. (2026). Sound in the multimodal construction of health and convenience: A case study of Sweetgreen's Instagram advertising. *Cognition, Communication, Discourse*, 32, 133-144.  
<https://doi.org/10.26565/2218-2926-2026-32-09>

### Abstract

Healthy food consumption and promotion are increasingly prominent trends in contemporary culture, reflected in the proliferation of healthy food products and their advertising in online media. Research has shown that such advertising is multimodal, yet sound is less frequently analyzed in detail than visuals and written language. The article aims to address this gap by demonstrating the meaning-making potential of sound in multimodal construction of health and convenience in a promotional Instagram Reel of a U.S. fast-casual restaurant Sweetgreen. Conducting a case study enabled us to perform a thorough analysis of sound and its interplay with the visual mode in order to highlight its potential importance as a semiotic resource in short-form promotional videos. The video was selected from a corpus of Sweetgreen Instagram Reels posted in 2025 as an information-rich example that foregrounds the presentation of a single product and relies on montage rather than spoken explanation or interview-style delivery, and combines music, object and ambient sounds. The visual track was analyzed through Kress and van Leeuwen's visual social semiotic framework and the auditory track through Graakjær's framework for sound and music to identify the meaning-making potentials of each track in isolation. Afterwards, the interaction between modes was examined through the adapted Martinec and Salway's system of status and logico-semantic relations to identify the contribution of each mode to the overall advertising message. The analysis shows that the visual montage compresses food preparation by omitting food labor and presents each ingredient as natural, pure, seasonal and fresh. The sound not only enhances the visual meanings, but also signifies the ongoing, coherent process of food preparation absent from the visual track, and frames it as artisanal craft rather than impersonal, mechanical action. The sound also uniquely provides circumstantial naturalness, thus adding a health-relevant meaning. Taken together, the sound is not merely decorative, but a key semiotic resource that helps convey brand values. Broader corpus-based analysis remains a task for future research.

**Keywords:** *multimodal analysis, case study, visual social semiotics, sound and music analysis, food advertising, Instagram Reel, healthy food branding.*

### 1. Introduction

Consumer trends in the early 21st century are defined by the rapid expansion of the global healthy food market, driven by rising diet-related diseases, ageing populations, and the increasing strain on national healthcare budgets (World Health Organization, 2025; Grand View Research, 2024). Research on healthy food advertising is driven not only by corporations seeking to capture a share of



this lucrative market (Grand View Research, 2024), but also by governments, as advertising can shape the concepts of health and dietary habits. A prominent example of such work is the EU-funded project “Infoodmation” launched in 2025. This initiative investigates the impact of mainstream media, social media, and direct marketing on consumer food choices, aiming to create governance frameworks that mitigate misinformation and foster transparent food communication (European Commission, 2024).

Advertising can be examined through various analytical lenses, including production, reception, experimental, and textual perspectives. From a textual perspective, an advertisement is treated as a text, and its analysis involves the identification of its meaning-making potential. Textual analysis attends to word choice, imagery, and other semiotic resources to identify both the messages audiences are explicitly invited to notice and the more implicit meanings that may lie outside the conscious awareness of producers and audiences. Such analysis yields valuable insights into how the advertisement works (Phillipov, 2021). Within the textual perspective, contemporary research often employs a multimodal approach to examining advertisements. Although human communication has always been characterized by multimodality, it gained a special prominence in light of the rapid spread of digitally-mediated communication (Weninger, Lim, & Chen, 2024). O’Halloran (2011) noted that the shift away from decoding the meanings of language alone occurred because of the need to account for various semiotic resources deployed in modern interactive digital technologies among other types of media.

The multimodal construction of healthfulness in food marketing has been examined across digital and print contexts. In online advertising, Koteyko and Nerlich (2007) applied a social semiotic framework to analyze how probiotic foods are promoted through a combination of verbal copy and images on websites. Chen and Eriksson (2019) used social semiotics and multimodal critical discourse analysis (MCDA) to show how corporate stories on healthy-snack websites integrate language, visuals, symbols, and colour to moralize “healthy eating.” Related MCDA work on organic food promotion likewise foregrounds visual and linguistic resources (Fernández-Vázquez, 2020). More recent platform research extends this to e-commerce: Yong and Rao (2024) analyzed Taobao “light food” ads using a textual-visual thematic framework on a dataset of images and accompanying text. Historical advertising research similarly applies a social semiotic framework to examine linguistic, typographical, visual, and compositional choices in health-oriented product promotion (O’Hagan & Eriksson, 2025). Across these healthy-food advertising studies, moving image and sound are less frequently analyzed explicitly than written language and static visuals. Ledin and Machin (2020) include a short social-media film clip in their Oatly case, but the bulk of the detailed analysis centers on packaging and poster campaigns. Where sound is foregrounded, it tends to be in narrower domains: sound symbolism in naming (Motoki et al., 2021) and sonic logos and their acoustic parameters in relation to perceived healthfulness (Techawachirakul, Pathak, & Calvert, 2022). Thus, while previous research has shown that food advertising is multimodal, the role of sound in the meaning-making of short-form healthy-food video advertising remains underdescribed.

Consumer focus on health brought to prominence fast-casual restaurants, which offered quick service, characteristic of fast-food restaurants, without compromising food quality (Krawczyk, Webb, & Mathe-Soulek, 2015). A U.S. fast-casual restaurant chain, Sweetgreen, explicitly builds its corporate story around resolving the perceived dichotomy between “fast” convenience and “slow” health. As part of that corporate story, the brand states:

- (1) *We started sweetgreen as three college students who were simply looking for a healthier way to eat. When we were in school, there were two choices: food that was slow, expensive, and fresh—or fast, cheap, and unhealthy. We saw an opportunity to create a business where quality was never sacrificed for convenience.* (Sweetgreen, n.d., “Our Story” section, para. 1).

This creates a semiotic challenge for the brand: the need to signify speed and operational efficiency without implying junk food and industrial processing.

This article *aims* to examine how visual and auditory resources jointly contribute to the construction of “healthy convenience” in Sweetgreen advertising, with particular attention to meanings supplied by sound.

The *objectives* of the study are to analyze the meaning-making potentials of the visual and auditory semiotic resources, to map the relations between them, and to determine the specific contribution of sound to the overall advertising message.

The *subject-matter* of the study is the multimodal construction of “healthy convenience” in digital food advertising.

The *material* of the study consists of a short-form promotional Instagram Reel published by Sweetgreen in 2025 and selected as an information-rich case for fine-grained multimodal analysis (Sweetgreen, 2025).

## 2. Methodology

This study employed a qualitative, single-case study design to examine the selected Sweetgreen Instagram Reel as a complex multimodal text comprising visual and auditory resources.

The specific restaurant, Sweetgreen, was chosen as the object of analysis because it is a highly visible brand with a broad footprint and an explicit corporate commitment to making healthy food compatible with fast-food convenience (Sweetgreen, Inc., 2025).

The Reel was selected from a corpus of 71 Sweetgreen Instagram Reels posted between January 1 and December 31, 2025. The selection criteria were as follows: (1) focus on the presentation of a single product; (2) montage-based organization rather than spoken explanation or interview-style delivery; and (3) presence of a soundtrack with music, object sounds, and ambient sound to allow for fine-grained multimodal analysis. The video was not treated as statistically representative of all fast-casual advertising. Rather, the Reel was approached as an information-rich and semiotically dense case whose highly controlled montage and sound design make the multimodal construction of “healthy convenience” especially legible. Thus, the research aimed at explanatory depth rather than corpus breadth.

Instagram was chosen as a platform that is frequently used by brands to market their products through audiovisual posts.

The analytical procedure was conducted in the following way. First, the video was deconstructed into its constituent semiotic tracks: the visual track and the auditory track. To isolate the specific contributions of each mode, the video was viewed multiple times in different sensory conditions: (1) without sound, to analyze the visual structure independently, and (2) without visuals, to analyze the auditory structure independently. The visual track was analyzed using Kress and van Leeuwen’s (2020) social semiotic framework, whereas the auditory track was analyzed using Graakjær’s (2014; 2021) framework for the analysis of sounds and music.

Finally, the relationship between the visual narrative and the categorized auditory layers was mapped using Martinec and Salway’s (2005) system for image-text relations. Although the authors’ classification system was originally developed for text and images, in the present study it was adapted to map out relations between moving image and sound. The status relations were determined through the assessment of the relative hierarchy of modes, based on rank scale and narrative completeness. The logico-semantic relations were then identified to specify how the auditory track related to the meanings that were expressed visually.

## 3. Findings and discussion

### 3.1. Visual track analysis

The video is a short montage of shots filmed against a continuous olive-colored cloth. It opens with white text on the screen (“sweetgreen,” “AUTUMN,” and the slogan “flavors to fall for”), followed by a disembodied hand that places a green paper bag with the inscription “sweetgreen” onto the surface. The montage then cycles rapidly through (i) utensils presented on a wooden cutting board

(fork, knife) and (ii) ingredients, which are shown either as whole produce placed on the cloth surface (apple, potato, Brussels sprouts, garlic, tomatoes, lettuce, Chinese cabbage) or as cut-in-half parts filmed against a white background (halved apple, halved Brussels sprout, sliced tomatoes). The ingredient shots include occasional minimal hand contact (placing, removing, touching). The sequence culminates in a reveal of the prepared salad in a bowl, marked by a shift in illumination. After that, the green paper bag is lifted away.

In the analysis that follows, each shot is treated as a still image for social-semiotic description in the framework of Kress and van Leeuwen (2020), and the montage is examined in terms of how it constructs a coherent visual discourse.



Figure 1. Ingredient shot staged on the olive cloth sweep (hand-tomato contact). Frame capture from Sweetgreen's *AUTUMN* [Instagram video].  
<https://www.instagram.com/reel/DP1cpNFjEPw/>



Figure 2. Ingredient "catalogue" insert against a white ground (tomato cross-section). Frame capture from Sweetgreen's *AUTUMN* [Instagram video].  
<https://www.instagram.com/reel/DP1cpNFjEPw/>

### 3.1.1. Representational meanings

Most ingredient shots across the sequence are organized as conceptual representations: single items and sets are presented as cleanly lit, isolated, and centrally placed, so that they seem to function as specimens rather than as parts of a narrated cooking process. This is reinforced by the catalogue logic of the sequence, where objects appear one-by-one against visually simplified grounds of the olive cloth surface and background (Figure 1) or the white background (Figure 2). Nevertheless, the sequence establishes a part-whole (analytical) relation via the Kuleshov effect, where the earlier shots with produce function as parts (possessive attributes) of the bowl of salad (the carrier) at the end of the video (Sperling et al., 2016).

At the same time, the video inserts narrative action processes via a disembodied hand. The hand places the branded sweetgreen bag onto the cloth-covered surface and later removes it, which signals the beginning and the end of the narrative. Another salient narrative moment occurs when the hand touches a single tomato with an index finger (Figure 1). The vector formed by the hand and its contact with the tomato construct an action process in which the tomato functions as the goal. This action can be interpreted as a gesture of inspection, which implies care in selection of ingredients and aligns with the brand's message to never sacrifice care for convenience. Additionally, the initially surprising absence of the food preparation processes could be a strategic choice aimed at showing the salad-preparation process as fast in order to signify convenience.

### 3.1.2. Interactive meanings

The interpersonal relationship constructed with the viewer could be classified as offer. Although the hand introduces human presence to the video, the viewers are not invited to engage with the implied person by, for example, a gesture. Presumably, what is expected of the viewers is to evaluate the quality of ingredients and the finished dish as observers. Close distance to the produce, its placement against the white background, and the tactile finger-to-tomato contact encourage the viewers to inspect the ingredients and to “feel” the freshness visually; it will be shown later that this is also supported by audio signification.

Angle and lighting further stabilize this inspection-oriented stance. The bowl is shown from a high angle perspective, which affords a comprehensive overview and encourages evaluative looking. Meanwhile, the illumination of objects against simplified grounds increases salience.

Finally, the alternation between the shots with the dimly lit, earthy background and the brighter, white background creates a modality mix: the former evokes artisan mood and seasonality and the latter evokes cleanliness, purity and analytical clarity. Together, these cues signal that the ingredients are traditionally grown and verifiably pure.

### 3.1.3. Compositional meanings

The title card uses vertical information value in a predictable Ideal-Real pattern: the brand name is placed at the top as the Ideal (brand identity) and the slogan at the bottom functions as the Real (the promised experience). The seasonal label “AUTUMN” is positioned centrally to anchor the reading of both brand identity and the promised experience within a seasonal frame.

Objects across the whole video sequence tend to be positioned centrally as well. At the same time, the surrounding field remains unoccupied. This compositional economy produces a minimalist aesthetic that may support perceived product naturalness (Saintives & Meral, 2024).

### 3.1.4. Montage-level meaning

Read as a sequence, the video constructs a compact visual narrative: seasonal announcement → the entry of branded packaging → the accumulation of cues that index craft and ingredient freshness and purity → the presentation of a finished bowl as narrative payoff → the branded packaging exits. Notably absent is the “messy middle” of food labor – washing, chopping, and mixing – whose omission is itself meaningful. Arguably, convenience is communicated by compressing temporally demanding material processes associated with healthy, “slow” food preparation into culturally legible signs of preparation and quality, that is, the knife and wooden cutting board, the staged whole ingredients, and moments of minimal inspection.

## 3.2. Audio track analysis

### 3.2.1. Reduced listening and iconic meaning potentials

Reduced listening focuses on the quality and structure of sound, which serves as a springboard for iconic signification. For analytic purposes, the soundtrack is divided into the diegetic sounds of knife, ingredients, touch, and birdsong and the non-diegetic music. The music uses a regular, repeated pattern (ostinato) that spans the whole audio track and could be transliterated as “ta-ta-TA-ta-ta,” where capitalization marks greater perceived prominence. The sound envelope of each percussive “ta/TA” is characterized by short attacks with no extended sustain. In terms of van Leeuwen’s parameters of sound modality, the dynamic and the pitch ranges are narrow with the prevalence of higher pitch sounds (van Leeuwen, 1999). However, an additional louder, lower-pitch pulse appears at the end of the soundtrack and functions as a cadential closure. There are no significant degrees of fluctuation or friction, and the absorption range is low, hence the music sounds rather “clean.”

Iconically, the temporal form of the music functions as a kinetic anaphone (Tagg, 2012). At first hearing, it seems to mimic the repetitive motion of automated machinery. Moreover, the regularity of the beat could act as an icon of industrial standardization. On the other hand, the regular, slight

variation in loudness, pitch, and notes as well as the cadential closure structurally resemble a well-organized procedure with stages and a marked completion point rather than monotonous motor movements. Arguably, this translates as a disciplined workflow at a well-run kitchen line.

The sounds of the bag, knife, touch, and produce seem hyperbolic and sometimes hyperreal: they are characterized by low friction, extremely close directionality, and low absorption range – a profile that could be associated with intensified advertising style, where objects sound their best. These close, “dry” object sounds could be interpreted as tactile anaphones that invite a “touch at a distance” sense (Tagg, 2012).

Finally, the birdsong introduces an ambient layer that creates the sense of perspectival depth. As the chirps appear at a slightly greater perceived distance than the close, “dry” object sounds and the music — which appear as “Figure” with no “Ground” — and come with a slight sense of reverb, they can iconically suggest a more expansive acoustic environment. This shifts the listening perspective from controlled workflow aesthetics to an airy, nature-coded setting.

### 3.2.2. Causal listening and indexical meaning potentials

Where in the mode of reduced listening, the focus is on sounds themselves, during causal listening, it is on the source of the sounds. Correspondingly, indexical signification relies on a direct causal connection whereby sounds inform the listener about material properties and interactions at a location in an environment. As a source-related classification, Graakjær’s tripartite distinction between the sounds “of,” “at,” and “with” food is especially useful for this type of listening (Graakjær, 2021). The music can be classified as a “sound with food” as an accompaniment that originates “outside” the food substance. The bag rustles and metallic clinks are “sounds at food,” as they emerge from the objects in the immediate vicinity of food preparation; and the handling of vegetables can be treated as “sounds of food” insofar as the audible texture is produced by the food substance itself.

The sounds of food are extremely important, as they can convey the material quality of the substance. The cells of fresh vegetables are characterized by high turgor, which contributes to hardness of the vegetable texture and produces higher-pitched sounds upon contact. Therefore, higher-pitched sounds index freshness and assist the listener in judging the quality of the food substance. This is exactly the type of sounds we can hear in the video, which supports the message that Sweetgreen never sacrifices quality for convenience.

The birdsong at the end of the soundtrack is also indexical in form. Simultaneous chirps and different timbres suggest multiple individuals, which – combined with the greater perceived distance and reverberations – constructs an impression that the sound comes from a large open space. The birdsong also indexes the time as birds normally chirp in the morning or at daytime. The placement of the birdsong as a background ambience, and the relative absence of competing urban noise, works indexically to situate the final salad reveal within a quiet, nature-coded environment. Since healthy foods are often positioned as “natural,” this supports the brand’s message about the healthfulness of ingredients (André, Chandon, & Haws, 2019).

### 3.2.3. Semantic listening and symbolic meaning potentials

Semantic listening asks what sounds refer to beyond their structural resemblance to other sounds and their physical source (Graakjær, 2021). This corresponds to symbolic relationship between signifier and signified, characterized by arbitrariness and social convention.

Research suggests that consumers perceive paper packaging as an attribute of healthy food and eco-friendliness (Sokolova, Krishna, & Döring, 2023). Moreover, paper packaging may connote artisanal craft, as a traditional packaging material. As a result, the sounds of paper may act as cues that implicitly convey health, eco-friendliness, and tradition by way of association (Granato, Fischer, & van Trijp, 2022).

Nature sounds, i.e., birdsong, have been shown to positively influence the willingness to buy organic foods. These foods are generally considered not only eco-friendly but also healthy, as they

are grown without the use of chemicals and do not contain genetically modified organisms (Spendrup, Hunter, & Isgren, 2016). Additionally, high-frequency twittering is associated with birds of a small size and metaphorically implies healthfulness, as healthy food usually comes in small portions (Motoki et al., 2021). Therefore, we can presume that birdsong may help to frame presented salad as healthy.

As regards the music, it features style indicators of classical music – the sounds of an acoustic piano, a bowed string instrument, ostinato, and cadential closure – which may evoke related cultural associations. There have been studies that associate classical music with preference for healthier and more expensive food options (Motoki et al., 2022; Tian et al., 2025; Areni & Kim, 1993). Presumably, classical music carries symbolic connotations that relate to food characterized by quality and healthfulness. Steady, clean ostinati recur in media contexts that privilege regulated continuity, such as process montages, controlled workflows, craft routines, as well as in contexts that cue tension or urgency (Lehman, 2012; 2017; McClelland, 2017). In the examined clip, the ostinato is anchored by a branded preparation-to-reveal montage and closed by a cadence at the finished bowl, making a controlled workflow field more congruent than suspense or threat. Combined with the symbolism of classical music, this reading aligns with Sweetgreen’s brand problem: it lets standardization be heard as practiced craft with repeatable quality rather than as industrial impersonality.

### 3.3. Multimodal analysis

#### 3.3.1. Status relations

Based on Barthes’ (1977a; 1977b) and Halliday’s (1994) classifications, Martin and Salway (2005) identified the following status relations between image and text: If the relation between image and text is unequal, then either the image is subordinate to the text, or the text is subordinate to the image. The subordinate component modifies the superordinate one. In case of subordinate image, the image relates only to a part of the text; in case of subordinate text there are two options – either the text relates to only a part of the image, or it must be decoded via image; for instance, the text may contain deixis that refers to the image. If, however, the relation between image and text is equal, the whole image relates to the whole text, and they can either complement each other or be independent. By analogy, the same categories will be applied to identify the relations between moving images and sound in Sweetgreen’s audiovisual text. Any continuous sound will be treated as the whole “text,” and the whole visual sequence – as the whole “image.”

The sounds of rustling, clinking, friction, and impact can be interpreted in a number of ways when they are separate from the visual sequence. For instance, the metallic clink could be the sound of drawing a sword from its sheath and the rustle could be produced by a person crumpling a sheet of paper. As we need the visuals to identify the exact participants indexed by the sounds, the sounds are not independent; they are “parasitic” on the image for referential specificity in Barthes’ terms (Barthes, 1977a). Arguably, the sounds are functionally similar to lexical deixis in that they index the processes in the shots. Analyzing the relationship of part versus whole, each sound is shot-specific and has no relation to all the processes across the clip; therefore, the status is unequal, and the sounds are subordinate to the visuals.

In the scene, where the view of the illuminated salad bowl is accompanied by birds’ chirping, the sound clearly modifies interpretation of the visual scene, implying an open window, fresh air, nature, and morning or daytime. It could also be argued that chirping constitutes an essential component, as it assists in interpreting the illumination as natural sunlight rather than as artificial lighting or as supernatural glowing added for effect. The visuals are also essential as they help identify bird’s chirping as ambience in a salad scene rather than, for instance, a wildlife documentary. However, this relation cannot be classified as complementary, because the birdsong and the visuals are tied via circumstantial information, not via processes (Martinec & Salway, 2005). Most importantly, the sound relates only to a part of visual sequence that constitutes the video – this qualifies the sound as subordinate.

Analyzing the music, it can be noted that it spans the entire clip and thus relates to the visual sequence as a whole rather than to a single isolated shot. As discussed in Section 3.2, the track is internally organized – a recurring ostinato with a cadential reinforcement at the reveal – and can be treated as a processual analogue that supplies a stable temporal frame across the montage. Although the visual sequence remains denotatively intelligible without sound, the music still modifies the sequence by supplying procedural continuity and circumstantial meanings. Conversely, the visuals anchor the otherwise “polysemous” ostinato in the domain of food preparation. The relation is therefore best described as equal-complementary, but with asymmetric anchoring, as the visuals carry the experiential nucleus.

### 3.3.2. Logico-semantic relations

To identify logico-semantic relations, we, first of all, need to identify participants, processes and circumstances in the video and audio tracks, as “*If the same participants, processes and circumstances are depicted and referred to, then there is elaboration. If new but related things are referred to or depicted, then there is extension. If related temporal, spatial or causal information is provided, then there is enhancement.*” (Martinec & Salway, 2005). In the shots with object sounds, the same participants, processes, and circumstances can be seen and heard across the modes. The sound and the visual events represent two realizations of the same micro-processes with the same level of generality, e.g., the thud of the Brussel sprout neither generalizes nor particularizes the depicted impact. Thus, the primary relation here is elaboration-exposition, where the sound re-expresses the same event via another semiotic channel. If one extends Martinec & Salway’s enhancement category, glossed as temporal, spatial, or causal, to include other circumstantial meanings, an enhancement-like reading becomes possible, in which the stylized object sounds contribute the manner cues of freshness. However, because these sounds are clearly designed and amplified, this remains a secondary interpretation rather than the primary classification.

In the salad reveal scene, the participants are the salad and the cloth-covered surface; the sudden light is a process; and the implied window to the outdoors is the circumstance. In the audio track, the implied birds are the participants, chirping is the process, and the implied outdoors is the circumstance. Since we have different participants, processes, and circumstances across the modes, the relation of elaboration can be excluded. Arguably, as the chirping introduces new participants (birds), the meaning represented by the visuals is extended by the sound event; however, functionally the chirping is not a new topic – it functions as an ambient setting for the salad shot. Therefore, the association of chirping with a natural setting can be deemed as added circumstantial information. Chirping also gives the viewer an idea about the time when the action in the video takes place as birds normally chirp in the morning or at daytime. Thus, we can classify the relation as enhancement, where the sound adds circumstantial information about when, where and in what type of environment the visual event takes place.

With respect to music, it can be divided into *Museme 1* and *Museme 2*. *Museme 1* accompanies the phase of salad preparation and gives a sense of ongoing, coherent process. Additionally, as it is carried by classical instruments, it conveys a sense of sophistication and cultural prestige, and thereby frames the preparation as controlled artisanal craft. *Museme 2* accompanies the final scene where the salad is presented to the viewers, and the closing shot where the branded bag is taken away. Musically, it is a cadence, which signals resolution. As both *Musemes* contribute circumstantial information – the manner of salad preparation in the case of *Museme 1* and the temporal endpoint in the case of *Museme 2* – the relations with the visual sequences can be classified as enhancement.

## 4. Conclusion

This article has shown that Sweetgreen’s short-form promotional video resolves the tension between health and speed not through explicit verbal claims, but through the interaction of visual and auditory resources. The visual track presents the ingredients as natural, pure, seasonal, and fresh, while also

compressing preparation by omitting food labor. This omission makes the soundtrack crucial, because it not only supports the visual meanings but also contributes meanings that are not fully inferable from the image sequence alone. Through ostinato, cadential closure, hyperreal object sounds, and ambient birdsong, the soundtrack signifies an ongoing and coherent preparation process, frames that process as controlled artisanal craft rather than impersonal mechanical action, and adds circumstantial naturalness. Sound therefore functions not as decoration, but as a key semiotic resource in the multimodal construction of “healthy convenience.”

Future research may test whether this pattern recurs across a broader corpus of fast-casual food advertising.

### Declaration of competing interests

The author declares that there is no conflict of interest or funding for this research.

The author declares no use of generative AI.

### References

- André, Q., Chandon, P., & Haws, K. (2019). Healthy through presence or absence, nature or science?: A framework for understanding front-of-package food claims. *Journal of Public Policy & Marketing*, 38(2), 172–191. <https://doi.org/10.1177/0743915618824332>
- Areni, C. S., & Kim, D. (1993). The influence of background music on shopping behavior: Classical versus top-forty music in a wine store. In L. McAlister & M. L. Rothschild (Eds.), *Advances in consumer research* (Vol. 20, pp. 336–340). Provo, UT: Association for Consumer Research.
- Barthes, R. (1977a). The photographic message. In S. Heath (Trans.), *Image, music, text* (pp. 15–31). London, England: Fontana. (Original work published 1961).
- Barthes, R. (1977b). Rhetoric of the image. In S. Heath (Trans.), *Image, music, text* (pp. 32–51). London, England: Fontana. (Original work published 1964).
- Chen, A., & Eriksson, G. (2019). The making of healthy and moral snacks: A multimodal critical discourse analysis of corporate storytelling. *Discourse, Context & Media*, 32, 100347. <https://doi.org/10.1016/j.dcm.2019.100347>
- De Temmerman, J., Vermeir, I., & Slabbinck, H. (2023). Eating out of paper versus plastic: The effect of packaging material on consumption. *Food Quality and Preference*, 112, 105023. <https://doi.org/10.1016/j.foodqual.2023.105023>
- European Commission. (2024). *Optimising food information and communication towards healthier and more sustainable dietary patterns* (Infoodmation; Grant agreement No. 101182813) [Project fact sheet]. CORDIS. <https://doi.org/10.3030/101182813>
- Fernández-Vázquez, J.-S. (2020). Selling organic candy: Multimodal critical discourse analysis of commercial websites. *British Food Journal*, 123(10), 3277–3292. <https://doi.org/10.1108/BFJ-09-2020-0800>
- Graakjær, N. J. (2014). *Analyzing music in advertising: Television commercials and consumer choice*. New York, NY: Routledge. <https://doi.org/10.4324/9781315770277>
- Graakjær, N. J. (2021). Analysing digital food sounds from a textual perspective: A case of champagne. In J. Leer & S. G. S. Krogager (Eds.), *Research methods in digital food studies* (pp. 40–53). London, England: Routledge.
- Granato, G., Fischer, A. R. H., & van Trijp, H. C. M. (2022). Meaningful reminders about the environmental impact of packaging: Effects of packaging cues on recycling intentions. *Journal of Environmental Psychology*, 79, 101724. <https://doi.org/10.1016/j.jenvp.2021.101724>
- Grand View Research. (2024). *Healthy foods market size, share & trends analysis report, 2024–*

2030. Retrieved April 14, 2026, from <https://www.grandviewresearch.com/industry-analysis/healthy-foods-market-report>
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London, England: Arnold.
- Koteyko, N., & Nerlich, B. (2007). Multimodal discourse analysis of probiotic web advertising. *The International Journal of Language Society and Culture*, 23, 20–31.
- Krawczyk, M., Webb, T., & Mathe-Soulek, K. (2015). The future of limited service restaurants: Let the consumer decide. *Journal of Hospitality & Tourism Cases*, 4(3), 11–20. <https://doi.org/10.1177/216499871500400302>
- Kress, G., & van Leeuwen, T. (2020). *Reading images: The grammar of visual design* (3rd ed.). London, England: Routledge.
- Ledin, P., & Machin, D. (2020). Replacing actual political activism with ethical shopping: The case of Oatly. *Discourse, Context & Media*, 34, 100344. <https://doi.org/10.1016/j.dcm.2019.100344>
- Lehman, F. (2012). Music theory through the lens of film. *Journal of Film Music*, 5(1–2), 177–196. <https://doi.org/10.1558/jfm.v5i1-2.179>
- Lehman, F. (2017). Manufacturing the epic score: Hans Zimmer and the sounds of significance. In S. C. Meyer (Ed.), *Music in epic film: Listening to spectacle* (pp. 27–55). New York, NY: Routledge.
- Martinec, R., & Salway, A. (2005). A system for image-text relations in new (and old) media. *Visual Communication*, 4(3), 337–371.
- McClelland, C. (2017). Of gods and monsters: Signification in Franz Waxman’s film score *Bride of Frankenstein*. *Journal of Film Music*, 7(1), 5–19. <https://doi.org/10.1558/jfm.27224>
- Motoki, K., Park, J., Pathak, A., & Spence, C. (2021). Constructing healthy food names: On the sound symbolism of healthy food. *Food Quality and Preference*, 90, 104157. <https://doi.org/10.1016/j.foodqual.2020.104157>
- Motoki, K., Takahashi, N., Velasco, C., & Spence, C. (2022). Is classical music sweeter than jazz? Crossmodal influences of background music and taste/flavour on healthy and indulgent food preferences. *Food Quality and Preference*, 96, 104380. <https://doi.org/10.1016/j.foodqual.2021.104380>
- O’Hagan, L. A., & Eriksson, G. (2025). Historicizing the marketing of plant-based meat substitutes: A multimodal analysis of Sanitas Nut Food Company advertisements, 1896–1901. *Food and Foodways*, 33(2), 103–130. <https://doi.org/10.1080/07409710.2025.2482483>
- O’Halloran, K. L. (2011). Multimodal discourse analysis. In K. Hyland & B. Paltridge (Eds.), *Continuum companion to discourse analysis* (pp. 120–137). London, England: Continuum.
- Phillipov, M. (2021). Textual analysis in digital food studies: New approaches to old methods. In J. Leer & S. G. S. Krogager (Eds.), *Research methods in digital food studies* (pp. 15–27). London, England: Routledge.
- Saintives, C., & Meral, H. (2024). Is it really natural? How minimalist food packaging influences consumers’ perception of product naturalness. *British Food Journal*, 126(11), 3888–3905. <https://doi.org/10.1108/BFJ-03-2024-0299>
- Sokolova, T., Krishna, A., & Döring, L. (2023). The perceived environmental friendliness bias in packaging. *Journal of Consumer Research*, 50(3), 468–491. <https://doi.org/10.1093/jcr/ucad008>
- Spendrup, S., Hunter, E., & Isgren, E. (2016). Exploring the relationship between nature sounds, connectedness to nature, mood, and willingness to buy sustainable food: A retail field experiment. *Appetite*, 100, 133–141. <https://doi.org/10.1016/j.appet.2016.02.007>

- Sperling, J., Barndt, K., & Kriebel, S. (2016, May 9). Montage. In *The Routledge encyclopedia of modernism*. London, England: Routledge. <https://doi.org/10.4324/9781135000356-REMO18-1>
- Sweetgreen, Inc. (2025). *Annual report pursuant to Section 13 or 15(d) of the Securities Exchange Act of 1934 (Form 10-K)*. U.S. Securities and Exchange Commission. Retrieved April 14, 2026, from <https://www.sec.gov/Archives/edgar/data/1477815/000162828025008367/sg-20241229.htm>
- Sweetgreen. (n.d.). *Our mission*. Retrieved April 16, 2026, from <https://www.sweetgreen.com/mission>
- Tagg, P. (2012). *Music's meanings: A modern musicology for non-musos*. New York, NY: The Mass Media Music Scholars' Press.
- Techawachirakul, M., Pathak, A., & Calvert, G. A. (2022). That sounds healthy! Audio and visual frequency differences in brand sound logos modify the perception of food healthfulness. *Food Quality and Preference*, 99, 104544. <https://doi.org/10.1016/j.foodqual.2022.104544>
- Van Leeuwen, T. (1999). *Speech, music, sound*. London, England: Macmillan.
- Weninger, C., Lim, F. V., & Chen, Y. (2024). Applying multimodal analysis: Embodied teaching and textbook analysis. In A. Riazi (Ed.), *Less frequently used research methodologies in applied linguistics* (pp. 68–82). Amsterdam, Netherlands: John Benjamins Publishing Company. <https://doi.org/10.1075/rmal.6.05wen>
- World Health Organization. (2025, September 25). *Noncommunicable diseases*. Retrieved April 14, 2026, from <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases>
- Yong, Q., & Rao, X. (2024). Exploring textual–visual strategies in internet-based light food advertising: A study of Taobao advertisements in China. *Humanities & Social Sciences Communications*, 11, 645. <https://doi.org/10.1057/s41599-024-03087-1>

### Sources for illustrations

- Sweetgreen. (2025, October 15). *AUTUMN* [Instagram video]. Instagram. Retrieved April 14, 2026, from <https://www.instagram.com/reel/DP1cpNFjEPw/>

## ЗВУК У МУЛЬТИМОДАЛЬНОМУ КОНСТРУЮВАННІ ОБРАЗУ ЗДОРОВОГО І ЗРУЧНОГО ХАРЧУВАННЯ: НА МАТЕРІАЛІ РЕКЛАМИ БРЕНДУ SWEETGREEN В ІНСТАГРАМІ

Андрій Ю. Рибальченко

Асистент,

Київський національний університет імені Тараса Шевченка

(вул. Володимирська, 60, м. Київ, 01033, Україна);

e-mail: [andrei.uni@live.com](mailto:andrei.uni@live.com)

ORCID: <https://orcid.org/0009-0007-1595-7429>

Рукопис надійшов 15.04.2026; отримано після доопрацювання 20.05.2026;  
прийнято до друку 25.05.2026; опубліковано онлайн 31.05.2026

### Анотація

Споживання здорової їжі та її популяризація дедалі виразніше утверджуються як тенденції сучасної культури, що виявляється у поширенні продуктів здорового харчування та їх реклами в мережі інтернет. Така реклама має мультимодальний характер, однак звуковий компонент значно рідше стає предметом докладного аналізу, ніж візуальний ряд і вербальний текст. Метою статті є дослідження ролі звуку у мультимодальному конструюванні брендівих цінностей «здоров'я» та «зручності» в рекламному відеоролику американського ресторану Sweetgreen в Інстаграмі. Дослідження окремого рекламного

кейсу дало змогу здійснити ґрунтовний аналіз звуку та його взаємодії з візуальним модусом і таким чином увиразнити його значущість як семіотичного ресурсу в короткоформатному рекламному відео. Відео обрано як інформаційно насичений приклад, увагу у якому зосереджено на презентації одного продукту, монтаж використано замість усного пояснення чи інтерв'ю, а звуковий ряд поєднує музику, звуки предметів та навколишнього середовища. Візуальний ряд проаналізовано за соціосеміотичною моделлю Кресса і ван Левена, аудіальний – за моделями аналізу звуку й музики Грок'єра для виявлення смислотворчого потенціалу кожного модусу окремо. Їхню взаємодію розглянуто за адаптованою системою статусних і логіко-семантичних відношень Мартинєка і Солвея, щоб визначити внесок кожного модусу в загальне рекламне повідомлення. Аналіз демонструє, що звук не лише підсилює візуально сконструйовані значення, а й репрезентує процес приготування їжі, відсутній у візуальному ряді, подаючи його як традиційну, а не індустріальну практику. Крім того, звук індексує природність, додаючи значення, пов'язане з концептом здоров'я. Отже, звук постає не просто декоративним елементом, а ключовим семіотичним ресурсом трансляції цінностей бренду. Ширший корпусний аналіз є завданням для майбутніх досліджень.

**Ключові слова:** *мультимодальний аналіз, соціальна семіотика, аналіз звуку й музики, кейс-стаді, реклама харчових продуктів, брендинг у сфері здорового харчування, відеоконтент в Інстаграмі.*

#### **Декларація про конфлікт інтересів**

Автор не має конфлікту інтересів щодо цієї статті.

Автор заявляє, що під час підготовки цієї статті не використовувалися інструменти штучного інтелекту.