Consumption Sounds: Vicarious Listening in Mukbang

"The open-mouth chewing makes my ears melt in satisfaction 🤤"

1. Introduction

The chewing, slurping, sipping and swallowing that accompany eating are ordinary bodily acoustics that rarely attract analytic attention. Yet in recent years, such sounds have been amplified, circulated, and aestheticised through the online phenomenon of mukbangs: livestreamed or recorded broadcasts in which audiences gather to watch, and crucially listen to, others eat. Here, the sounds of eating are not incidental. They are part of the attraction. While the visual display of food matters, and while images help viewers recognise what they are hearing, this paper focuses on the auditory dimension of mukbang. It asks how ingestion sound helps create a sense of eating with another person at a distance.

Patterson and Larsen (2019) urge researchers to listen to consumption and to treat sound as an analytical lens rather than as background. Mukbang offers a particularly direct case for examining consumption through sound because it places the audible act of eating at the centre of the format itself. The sounds of eating become part of what is staged, sought out, and shared. At the same time, these sounds do not operate alone. They are heard through visible food, mouths, microphones, camera proximity, titles, tags and platform conventions. This makes mukbang a useful site for asking how consumption becomes audible within a mediated consumer culture format.

"*The crunch gods are blessing me tonight* 😍"

Existing mukbang research has linked the format to companionship, co-presence, commensality, parasocial closeness, vicarious satisfaction and consumer intention (Anjani et al., 2020; Aucoin, 2019; Choe, 2019; Ferdous et al., 2016; Sanskriti et al., 2023; Song, 2021, 2023; Wang et al., 2021). The present paper focuses that discussion on the auditory organisation of eating. It develops the concept of *vicarious listening* to describe a mediated relation in which another person's consumption becomes sonically available. The central question is how ingestion sound contributes to *mediated accompaniment* by giving eating the form of an unfolding bodily action within the video format.

The study uses repeated close listening to the original audiovisual clips, supported by attention to platform cues such as format, tagging, and visible recording conventions, to analyse publicly available mukbang videos. It offers an interpretive, not experimental, account. The claims apply only where ingestion sounds are proximate and temporally continuous. Highly edited, music-dominated, or primarily visual formats fall outside the argument. Within those limits, the paper shows how bodily consumption becomes part of mediated consumer experience through its auditory presentation.

2. Literature Review: Mukbang, Listening and Mediated Consumption

2.1 Mukbang as mediated commensality

Mukbang foregrounds ingestion sound in a particularly direct way. It is also a form of mediated consumption in which food, bodies, screens, audiences and platforms are organised around the

act of eating at a distance. Prior work explains why mukbang may feel socially and sensorially meaningful (Anjani et al., 2020; Aucoin, 2019; Choe, 2019; Ferdous et al., 2016). Some viewers use mukbang to accompany solitary meals, relieve loneliness, experience closeness, relax, or enjoy the spectacle of another person's eating (Anjani et al., 2020; Kang et al., 2020; Song, 2021, 2023).

Choe's (2019) interactional analysis is especially important because it shows that eating sounds are coordinated with talk, gaze, and viewer prompts as part of a shared event. Mukbang is therefore multimodal from the start. Eating is shown, narrated, responded to and timed for an audience. Yet within this literature, sound is usually treated as one element of the interaction, with less attention to its own organisation. The present paper builds on Choe's multimodal account by shifting the emphasis toward ingestion sound itself. It examines how eating is given an audible sequence: a bite begins the action, chewing sustains it, a swallow closes the moment, and a pause or breath prepares the next mouthful. The claim is specific: the audible sequence gives attention something to stay with as the meal unfolds.

Therefore, this paper examines how eating sounds help sustain a sense of accompaniment in mukbang. In some livestreamed mukbangs, viewers and hosts interact directly through comments, prompts and responses. In many recorded or replayed clips, however, the relation is less reciprocal. Even there, the clip can guide attention through repeated patterns of bite, chew, swallow, breath, pause and return. The issue is how sound gives that relation its temporal shape.

"I am eating Korean fried chicken along with you right now."

2.2. Listening to consumption

This matters in part because consumer research has more often analysed consumption through what can be seen than through what can be heard (Kavanagh, 2013; Patterson & Larsen, 2019). Work on atmospherics and music has long shown that sound can shape behaviour and evaluation, but it is often treated as a functional stimulus within a broader setting rather than as an analytic object in its own right (Garlin & Owen, 2006; Turley & Milliman, 2000). Bradshaw and Holbrook's (2008) critique of Muzak offers an important consumer-culture precursor because it treats marketplace sound as a matter of power, control and cultural experience rather than as a neutral atmospheric variable. Patterson and Larsen (2019) push further by asking scholars to listen to markets and consumption themselves. Their intervention helps shift the question from what sound does in the background to how sound participates in the organisation of consumption.

"I love ur eating sounds 😊😊😊*"*

Sound Studies helps clarify this further. Listening unfolds in time, is shaped by position and embodiment, and can treat sound as a way of knowing rather than as mere signal detection (Feld, 2015; Medina, 2023). Patterson and Larsen's (2019) sonic turn is especially useful here because mukbang is not simply a case where sound influences consumption from the outside. It is a format in which consumption itself sounds. This is particularly important for mukbang because the small bodily actions of eating are presented as an unfolding sequence of sounds. A listening-led approach therefore directs attention to how eating becomes audible over time

rather than to visual display alone. Recent sonic-turn work extends this agenda by asking how different forms of consumption become available through particular listening practices, devices and sonic materials (Ryan & Ruiz, 2026). Mukbang adds a bodily and performative case to that discussion. It stages ingestion as something that can be heard.

The listening approach needs precision. Chion and Gorbman's (1994) distinction between causal and reduced listening is useful because it separates listening for a sound's source from listening to its sonic qualities. In the present analysis, much of the listening is causal: sounds are interpreted in relation to recognisable bodily and food-related sources within the eating scene. Smith's (2025) distinction between listening-reading and listening-responding is also useful because mukbang sounds are interpreted within a visible and culturally recognisable eating scene. The analysis therefore treats mukbang listening primarily as causal listening and listening-reading, while using reduced listening more modestly to attend to qualities such as wetness, grain and density. Because the study analyses publicly available videos rather than viewers' embodied responses, Smith's distinction is used mainly to clarify that the analysis is centred on listening-reading rather than listening-responding.

This formulation keeps sound at the centre without isolating it from the video format. Mukbang sounds are heard through visible food, mouths, gestures, microphones, titles, tags and platform conventions. Eating sound rarely arrives without a recognisable source. The contribution is to examine how mukbang arranges ingestion sound as part of the consumer experience, giving bodily eating a sonic form that can be interpreted within the video.

2.3. Ingestion sound, vicarious listening and formatted audibility

Evidence from food and flavour studies further justifies attention to ingestion sound. Altering bite acoustics changes perceived crispness and freshness even when the food itself is unchanged, while background music and ambient noise can shift judgements of creaminess, bitterness, and liking (Reinoso Carvalho et al., 2017; Spence, 2015; Spence & Wang, 2015; Zampini & Spence, 2004). The paper does not retest those causal findings. It draws on them only to support the more specific point that sound matters to eating-related experience, making mukbang a plausible site for listening-led analysis (Anjani et al., 2020). Mukbang intensifies this audibility by placing ingestion sound close to the microphone and at the centre of attention.

The term vicarious, however, requires care. In one sense, the existing literature points toward mediated commensality: viewers may feel that they are eating with someone at a distance (Choe, 2019; Ferdous et al., 2016). In another sense, some studies suggest something closer to proxy gratification. Anjani et al. (2020), for example, report vicarious pleasure and virtual satiation among some viewers, including cases where chewing sounds make it feel as though they are doing the eating themselves. Related work also describes second-hand eating, craving relief, and surrogate satisfaction as part of mukbang viewing (Song, 2021). These findings are useful, but they also mark an important boundary for the present study. This paper examines how ingestion sound is organised in these videos, rather than measuring appetite, satiation, craving reduction or behavioural outcomes. Used in this way, vicarious listening helps explain how a sense of accompaniment can be organised through audible ingestion, while leaving open the more expansive question of whether some viewers also experience something closer to proxy consumption.

The concept overlaps with ASMR, and some mukbang clips are explicitly tagged as such. ASMR research typically foregrounds tingling, calm, or relaxation across a broad range of delicate sounds (Poerio et al., 2018). For the present argument, its usefulness lies in showing how close sound, whispering, bodily detail and platform address can create a sense of mediated nearness (Zappavigna, 2023). This also means that vicarious listening cannot be treated as a purely subjective response detached from mediation. Microphones, camera placement, food choice, editing, creator routine and platform genre help determine which sounds become close, clear, intense or acceptable. These are not neutral technical details. They affect how eating becomes available as an object of listening in the first place (Sterne, 2012). Breathing, chewing, wetness and swallowing can be amplified, softened, segmented or made continuous. Vicarious listening is therefore approached here as dependent on *formatted audibility*: the produced sonic conditions through which ingestion is made available for sustained listening. These conditions include proximity, so that mouth and food sounds appear close enough to register as bodily; continuity, so that eating can be followed as an unfolding sequence; pacing, so that bites, chews, swallows and pauses become rhythmically organised; and texture, so that the qualities of food and mouth contact become perceptible. These conditions give ingestion a patterned sonic form within the mukbang format, adding an unfolding auditory layer to the visual display of eating.

3. Analytic Framework and Methods

The aim of this section is transparency, procedural clarity, and auditability: a reader should be able to understand how clips were selected, how listening was conducted, and how the interpretations were reached. The approach is interpretive rather than experimental. To support

consistent description across cases, the analysis uses a six-element listening lens derived from the paper's account of vicarious listening and from recurring audible features identified during repeated listening across the selected mukbang clips. The lens supports systematic description rather than causal inference. It combines source recognition, audiovisual interpretation, and attention to sonic qualities, allowing the analysis to move between what produces the sound, how it is situated in the video, and how it is heard as texture, proximity, pacing and continuity. Each element isolates one feature of the clip's sonic organisation.

3.1. Six-Element Listening Lens

The six elements are pacing, proximity, continuity, onsets and resets, timbre and wetness, and address and turn cues.

Pacing refers to the recurring pattern created by bites, chews, swallows, and brief quiet intervals. The analytic question here is whether ingestion unfolds with enough regularity of tempo and phrasing to be followed over time. Pacing is used here in an ordinary temporal sense, rather than as a technical musical term.

Proximity concerns how near the recording places the listener to the eater's mouth and breath. Close capture, high recording levels, and reduced room noise can make small details such as breath, swallows, and tongue contact more audible than they would be in everyday face-to-face listening. The analytic focus is on how sonic closeness is produced within the recording.

Continuity refers to the stability of the audio flow across time. The question is whether the ingestion sequence remains intact or is broken by jump cuts, heavy background music, or

extended talk that disrupts its temporal pattern. This element is useful for marking both the presence and the limits of the relation described in the paper.

Onsets and resets refer to the clarity of audible starts, such as bites or slurps, and the brief pauses that allow each event to register distinctly. The focus here is on how one ingestion event is separated from the next, and how these moments contribute to the pacing of the sequence. In this limited sense, resets also work as a form of temporal punctuation: they help one bite, swallow, sip or breath register before the next event begins (Albert & Bell, 2002; Chion & Gorbman, 1994).

Timbre and wetness refer to the acoustic character of the eating sounds themselves, including whether they register as crisp, soft, dense, dry, or wet, and how these qualities are balanced in the mix. Here, wetness refers to audible moisture in chewing, swallowing, lip contact, or short suck-in sounds. The element attends to the texture of the sounds rather than to their mere presence. It is therefore the element most directly concerned with sonic quality rather than source identification alone.

Address and turn cues refer to small signals that help organise anticipation and timing, such as a breath before a bite, a sip following a spoken remark, a pause that invites response, or brief verbal address aligned with ingestion. The focus is on how speech, breath, and ingestion are coordinated within the interaction. I use “turn cues” descriptively, not as a formal conversation-analytic category. The term refers to moments where sound helps mark readiness, response, transition or renewed eating.

3.2. Corpus and Sampling

I built a purposive set of mukbang videos in which eating sounds are central to the performance. Items were included when they (a) foregrounded chewing, sipping, swallowing, or utensil contact, (b) contained sustained stretches in which ingestion was not masked, and (c) showed some platform-level indication that sound mattered, through titles, tags, descriptions or viewer comments. These criteria align with prior work describing mukbang as a multimodal eating broadcast in which eating sounds, camera-facing performance and viewer interaction form part of the genre (Choe, 2019; Kang et al., 2020).

The selected clips were drawn from YouTube and included both long-form and short formats. For each item, I recorded basic metadata, including creator handle, platform, date, duration, creator-supplied tags, and any disclosed recording notes. The final set comprised 24 videos (N = 24), with a total running time of 10 hours and 42 minutes, collected between January 2024 and June 2025. The sample included 14 Korean-language videos, 8 English-language videos, and 2 in other languages, spanning Korean, U.S. fast food, Japanese, Chinese, and fusion food contexts. The inclusion of Korean-language videos reflects the Korean origins and continuing prominence of mukbang as a platform genre. The analysis does not generalise about Korean eating culture. Short viewer comments quoted in the paper were drawn from the selected mukbang videos and used selectively as contextual material when they explicitly referred to eating sounds, listening, sonic pleasure, discomfort, or eating along with the creator. They were not treated as a separate audience-response dataset.

The sampling sought variation rather than representativeness. I included sound-led and talk-led formats, channels tagged as ASMR-adjacent and channels without such tags, quieter and

busier ambient settings, and creators eating alone as well as with a companion on screen. I screened 54 items in total and retained 24. Common reasons for exclusion included heavy background music, extensive jump cuts that removed ingestion sequences, poor audio capture, capture distance, audible automatic gain control, and persistent masking that made ingestion difficult to follow (Kang et al., 2020). In that sense, the framework applies only where eating is deliberately made audible and available for sustained listening.

3.3. Analytic Frame and Listening Procedure

All included items were analysed using the six-element listening lens introduced above so that descriptions remained comparable across cases. The three listening approaches were operationalised according to what each made legible or, in the case of reduced listening, what it temporarily set aside. Causal listening was used to make sound sources legible by identifying chewing, swallowing, breathing, sipping, utensil contact and packaging sounds. Listening-reading was used to make the audiovisual situation legible by interpreting those sounds in relation to visible food, mouth movement, gesture, apparent recording proximity, title, tag and interactional context. Reduced listening was used more narrowly to set aside source recognition and audiovisual context as far as possible, and to attend to sonic qualities less directly tied to source or contextual legibility, such as density, wetness, grain, sharpness, softness, pause and pacing. Choe's (2019) interactional analysis also supports treating eating sounds as part of how mukbang scenes are organised, rather than as incidental background noise. I use the six-element framework descriptively, as a means of structured comparison rather than causal inference.

I applied the same four listening passes to every item. In the first pass, I logged the overall sonic scene and noted recurring sound patterns. I also noted the visible setting, camera position,

food type, creator address and platform cues, so that listening was situated within the audiovisual format. In the second pass, I mapped the ingestion sequence through time-coded notes for bites, chews, sips, swallows, breaths, utensil or surface contact, packaging sounds, pauses and quiet. This pass relied mainly on causal listening because it identified what produced the sound and where it occurred in the sequence. In the third pass, I used listening-reading to interpret those sounds in relation to on-screen action, social address and platform cues. In the fourth pass, I listened without viewing the image and attended to the audio as sound, focusing on qualities such as grain, density, sharpness, wetness, softness, pause and pacing. This final pass did not replace the audiovisual analysis; it clarified which sonic qualities remained noticeable when the image was withheld and contextual cues were set aside as far as possible.

Time-coding served three purposes. It made specific moments auditable, supported comparison across clips, and helped mark the limits of the framework by showing where continuity held and where masking or editing interrupted it. I also kept notes on ambiguous cases, such as mouth air versus room air, or masking versus intentional quiet. When an item consistently failed the scope conditions, for example because ingestion audio remained masked, it was logged as an exclusion with boundary notes and removed from cross-case comparison.

3.4. Comparison and Audit

For each included item, I identified at least three signature, time-coded moments and recorded four features: the timecode, the relevant analytic element, a descriptive note on sonic features, and an interpretive note on how the moment contributed to temporal flow, accompaniment or recording nearness. A short narrative summary then described the overall tempo, phrasing, and

use of address. These records were kept in a spreadsheet and supplemented with short memos where judgement calls affected inclusion, exclusion, or interpretation.

Using this grid, I compared cases across different creators and formats, including sound-led clips, talk-led clips, and clips where editing shaped the continuity of ingestion sound, in order to specify where vicarious listening applies and where it weakens. A second listener audited approximately 25% of the included items (n = 6) to check the salience of coded moments. Discrepancies were discussed and recorded in short decision notes. The aim was not to eliminate interpretation, since the approach is explicitly interpretive, but to check for possible interpretive bias.

3.5. Researcher Stance and Ethics

Because ingestion audio can be intimate or unpleasant, the analysis required attention to researcher stance. I conducted the study as a researcher familiar with sound-led media and attentive to cultural etiquette around eating sounds. I also monitored for misophonia sensitivity during the research process and planned breaks where needed. When the sounds produced attraction or aversion, I treated those responses as information about audibility and social norms, not as judgements about the creators themselves.

The study used publicly available mukbang videos on YouTube and did not involve interaction with creators or viewers. I followed platform terms and standard internet-research guidance. Channels were de-identified where feasible, and comment excerpts were curated with care, especially where threads included harassment or discussion of disordered eating. The focus of the analysis is on how sound is produced and organised, not on making claims about health or

about particular individuals. The study was treated as low-risk internet research based on publicly available material.

4. Mukbang Cases

This section applies the six-element listening lens to three short cases. Each case focuses on a small number of signature, time-coded moments where the organisation of eating sound is especially clear. The aim is not to document every audible event, but to show how the listening lens can be used to produce a disciplined comparison across clips. The three cases were selected because they make the listening lens visible in different sonic arrangements.

The supplementary audio files correspond to the three case excerpts discussed below. They are provided so that readers can listen with the analysis, especially to the sonic qualities described in the time-coded moments. They do not replace the original audiovisual videos on which the analysis is based, since the paper treats mukbang sound as heard within a visual and platformed format. Their value is more specific: by temporarily separating sound from image, they allow readers to attend more closely to grain, attack, density, wetness, pause and pacing, and to consider what becomes noticeable when the sounds are heard apart from the visible eating scene.

4.1. Case A (insert audio file): Close-mic crunchy food

This is an audio-led, close-mic recording built around a crunchy fried food. Bright bite onsets, softer damped chewing, and brief pauses create a clear temporal pattern. The food source is

visually recognisable, but the clip gives particular emphasis to the contrast between the sharp bite sounds and the softer chewing that follows.

Signature moments

00:00. Onsets and resets; timbre and wetness

A double crunch arrives in quick succession, with the first onset slightly louder and followed by a long decay. This establishes a bright, nearfield texture from the start. The onset is heard, in relation to the visible eating action, as a bite, while the sonic quality is defined by sharp attack, brittle breakage and close capture.

00:05 to 00:07. Pacing

The decay resolves into a steady closed-mouth chewing rhythm. The sequence becomes regular enough to be followed over time.

00:21 to 00:27. Onsets and resets; continuity

A brief pause is followed by a slower-attack bite with a long decay, then open-mouth chewing. The short reset makes the next onset distinct and helps preserve the continuity of the sequence. The pause does not interrupt the eating event. It separates one bite from the next and gives the following onset clearer definition.

00:42 to 00:51. Proximity; continuity

The clip alternates between open-mouth and closed-mouth chewing. At 00:51, a brief quiet interval contains an audible swallow followed immediately by a new bite. The swallow-reset makes small bodily transitions clearly audible and keeps ingestion in the foreground. Here, the sequence can be followed as a transition from chewing to swallowing, while close capture makes that transition sonically prominent rather than merely visible.

01:14. Proximity; address and turn cues

A short sucking intake before resumed chewing places the listener unusually close to the

action and works as a nonverbal cue for the next move in the sequence. The intake functions as a small anticipatory cue: it prepares the resumption of chewing without requiring speech.

Synthesis

Case A is structured by bright crunch onsets, softer chewing phases, and short audible resets. The clip demonstrates how a stable pacing can be sustained through clear attacks, brief pauses, and nearfield detail. Proximity makes mouth-state shifts and swallows audible, while small intake sounds help cue the next phase of the sequence. Together, these features create a clear example of ingestion made followable through crunchy, high-contrast audio. The case shows how source recognition and sonic quality work together: bites, chews and swallows remain identifiable, while sharpness, decay, dampening and pauses give those actions a distinct temporal form.

4.2. Case B (insert audio file) Wet slurp-chew-swallow sequence

Unlike Case A, this clip is organised around recurring slurp-chew-swallow patterns. The microphone is close, the room is quiet, and there is no speech or music, so very small mouth events remain audible throughout. The clip is defined less by sharp attack than by wet, repeating phrasing.

Signature moments

00:00 to 00:06. Pacing; continuity

A long ingressive slurp leads into closed-mouth wet chewing, followed by a brief quiet and a clear swallow at 00:06, then immediate re-intake. The clip establishes its basic slurp-chew-

swallow cycle immediately. The sequence can be followed as intake, chewing and swallow, while the close recording makes the transitions between these moments unusually prominent.

00:09 to 00:11. Proximity; timbre and wetness

Closed-mouth wet chewing shifts into brighter open-mouth chewing. Heard in relation to the visible mouth action, this change registers as a shift in mouth position because the microphone captures fine changes in wetness and tone.

00:19 and 00:22. Continuity; pacing

Two swallows are folded into the ongoing rhythm without breaking the sequence. This maintains a stable temporal flow and keeps the pattern easy to follow.

00:38 to 00:40. Address and turn cues; proximity

A soft breath is followed by a firm sucking intake. These bodily cues work as nonverbal signals that prepare and mark the next cycle. They also show how breath is integrated into the eating sequence rather than sitting outside it.

01:55 to 02:13. Onsets and resets; timbre and wetness

A brief quiet interval is followed by a slower intake at 01:58, renewed wet chewing at 02:00, brighter open-mouth wetness at 02:08, and a swallow with breath at 02:13. The reset and return sharpen attention to the clip's wet, intimate acoustic character. Here the pause does not stop the sequence. It clears space for the next intake and makes the renewed wetness more noticeable.

Synthesis

Case B shows a more fluid form of patterned ingestion than Case A. Its core structure is not a sequence of isolated attacks but a repeated slurp-chew-swallow unit in which breath is often folded into the intake itself. Proximity makes throat transfer, lip contact, and small shifts in mouth opening audible, while brief quiet intervals pace the sequence without breaking it. The

result is a more fluid and intimate version of the pattern than in Case A, but one that remains equally dependent on continuity and followable phrasing.

4.3. Case C (insert audio file) Breath-intake sequence

This is a close-mic, audio-led clip in which breath and intake are tightly coupled. Cooling blows, nasal exhales, and lip smacks punctuate the sequence, and a mid-clip shift to a harder ingredient introduces a temporary crisp contrast within an otherwise wet texture. No speech or music is present.

Signature moments

00:02. Pacing; address and turn cues

A loud ingressive slurp is followed by four timed chews aligned with a two-part nasal exhale. Breath and chewing are coordinated strongly enough to establish a clear rhythmic unit from the outset. The sequence is heard, in relation to the visible eating action, as slurp, chew and exhale, while the timing of the breath gives the unit its particular shape.

00:07 to 00:10. Continuity; proximity

An immediate re-slurp is followed by audible intra-oral movement, with breath released at 00:10. The close capture makes small mouth movements audible as part of the sequence without breaking its continuity.

00:13. Address and turn cues

A rhythmic nasal exhale, a light smack, and another intake work together as a nonverbal cue for the next phase of the clip. The cue is bodily rather than spoken: it marks transition through breath and mouth sound.

00:21 to 00:22. Timbre and wetness

A harder ingredient enters, producing a dry crunch before the sound folds back into wet chewing. This briefly alters the acoustic profile and refreshes attention through contrast.

00:29 to 00:34. Onsets and resets; continuity

A deep inhale, exhale, and double cooling blow prepare the next module without masking the mouth sounds. These preparatory respiratory sounds function as pacing devices rather than interruptions. They reset the sequence while keeping ingestion audible.

Synthesis

Case C demonstrates a form of patterned ingestion in which respiration is integrated into the sequence itself. Breath, cooling blows, and lip smacks actively organise phrasing, while the brief introduction of a harder ingredient adds contrast within a predominantly wet timbre. The case shows how the pattern can be sustained even when the pacing depends less on simple chew cycles and more on tightly coupled breath-intake transitions. Compared with Cases A and B, this clip makes respiratory sound especially important to followability: breath helps mark when one phase ends and the next begins.

5. Discussion

The cases clarify the sonic work involved in mediated accompaniment. Across the three clips, recorded ingestion is given temporal shape: bites begin moments, chewing sustains them, swallows close them, and brief pauses or breaths prepare their return. Close recording adds bodily nearness by bringing breath, saliva and mouth-state changes into the foreground.

5.1 What the cases clarify

Across the three cases, the most stable finding is temporal organisation. Case A shows a high-contrast pattern in which crisp onsets and short resets make the sequence easy to track. Case B achieves a similar effect through repeating slurp-chew-swallow units. Case C shows that a comparable followable structure can be produced through breath phrasing, cooling blows, and tight breath-intake transitions, even when chewing is not the only unit. The sonic textures differ across the cases, but the temporal structure remains consistent.

Mukbang divides eating into recognisable and anticipated units. Pacing gives the sequence temporal form, resets separate one event from the next, and continuity keeps the sequence intact so that the listener can remain with it. Nearness matters because it turns small bodily events into reliable cues. Close capture brings breath changes, swallows, shifts in mouth state, and minor intake sounds into the foreground. When recording distance increases, when masking enters, or when editing breaks the flow, the sequence becomes less trackable.

This nearness is produced through the format. Recording proximity and sound balance position the listener unusually close to bodily processes that are ordinarily socially buffered. The result is intimacy without shared physical space, smell, touch, or reciprocal obligation, but with breath and saliva made acoustically present. The relevant relation is therefore temporal as well as sensory. The format makes it possible for attention and anticipation to align with the eater's rhythm, even though the present analysis does not measure whether viewers actually mirror that rhythm.

The cases also clarify the boundary between eating with and eating through. The broader mukbang literature leaves open the possibility that some viewers may experience proxy gratification, virtual satiation, or second-hand eating. The present material supports a more bounded claim: recorded ingestion sound can help organise a listening position in which accompaniment becomes plausible.

5.2 Theoretical contribution

The paper makes three main theoretical contributions. First, it specifies what "listening to consumption" can mean in a concrete consumer setting. Following Patterson and Larsen (2019), the cases show that consumption can be studied through its temporal arrangement as sound. In mukbang, listening draws attention to the moment-by-moment organisation of eating: the start of a bite, the continuation of chewing, the closure of a swallow, and the preparatory breath or pause.

Second, the paper clarifies what "the sound of consumption" refers to in this case. The focus is a bodily register made central and repeatable. Formatted audibility names the condition in which these bodily sounds are stabilised as a repeatable object of attention. That stabilisation matters because it supports a repeatable sensory experience within a platformed format. The food remains visible, but the consumable object expands. What is offered is food, mouth, microphone, timing and texture arranged as a listening position.

This also helps differentiate registers of consumption sound. One register is infrastructural: systems that hum, regulate, transmit, and govern consumption without directly addressing us (Ryan & Ruiz, 2026). Another is bodily and performative: mouths, breaths, swallows, and

textures. Mukbang pushes this bodily register to the foreground, making ingestion sound part of the consumer offering rather than a background accompaniment. The question of how such bodily sound becomes acceptable, desirable and shareable is developed in the next section through the idea of sonic etiquette.

Third, the paper defines vicarious listening as a way of following another person's consumption through sound. Through repetition, reinforced boundaries, and anticipated sequencing, the format makes crunch, wetness, breath, and swallow available as meaningful cues within a patterned flow. In this formulation, formatted audibility names the produced sonic conditions, while vicarious listening names the relation those conditions make possible. This keeps the concept distinct from proxy consumption: it names a mediated relation to another person's eating, not a claim that the viewer literally consumes through the performer.

This also clarifies the two-way character of mediation in the paper. Sounds are not treated as autonomous objects: they are heard within an audiovisual and platformed scene that helps make their sources and contexts legible. At the same time, the eating event is not merely illustrated by sound. Audible ingestion gives the visible eating event temporal contour and helps organise the sense of mediated accompaniment. Mukbang therefore works through a double movement: the format helps listeners know what they are hearing, while sound helps organise eating as something that can be followed.

The broader theoretical implication, developed below, is that consumer intimacy can be organised through the formatting of bodily sound. Mukbang may create closeness without shared physical space, sensory involvement without shared food, and accompaniment without reciprocal obligation. This relation is managed through timing and distance. The listener is

brought close to the eater's mouth, but only through a format that controls what kind of closeness is possible. This is why mukbang matters for consumer culture research: it shows how platforms can turn the ordinarily buffered sounds of bodily consumption into a mediated sensory relation.

5.3 Bodily sound as platformed consumption

The wider consumer culture implication of mukbang is that bodily ingestion becomes part of what is offered. Existing mukbang research has shown how online eating can support mediated commensality, companionship, vicarious pleasure and multisensory experience (Anjani et al., 2020; Choe, 2019). This paper adds a more specific claim: those possibilities are partly organised through the formatting of ingestion sound. The cases show how patterned audibility gives different roles to crunch, wetness and breath. The paper therefore shifts attention from the online display of eating to the way bodily consumption is recorded and arranged as an event that can be heard while it unfolds.

At this level, listening to consumption means attending to how bodily ingestion is organised through recorded sequence and proximity. Across the cases, crunch, wetness and breath provide different routes into the same arrangement: the mouth becomes more than the place where food disappears. It becomes the listener's point of access to the consumption event, where food, microphone and attention converge.

This is the sense in which the mouth becomes an interface. Within the format, the mouth becomes the point through which the consumption event is sequenced for others. A bite opens an event, chewing sustains it, the swallow closes it, and breath or pause prepares its return. In

ordinary eating, these sounds often recede into the background of shared social life. In mukbang, they are moved forward and given structure. The video presents food being eaten and also gives access to the small transitions through which eating happens.

The cases also show how mukbang repositions mouth sounds within the format. Audible chewing, slurping, swallowing and saliva have an unstable social status. In many face-to-face meals, they are managed as sounds to be softened, ignored or kept within limits. Mukbang gives these same sounds a different role. Case A makes crunch attractive through sharpness and clean separation. Case B makes wetness sustainable through repetition and continuity. Case C incorporates breath into the sequence instead of treating it as accidental background. The format therefore creates what might be called a *sonic etiquette of eating*, extending the idea of sonic etiquette into the recording and platforming of bodily consumption (Kytö, 2018). In ordinary table manners, etiquette often concerns whether mouth sounds should be audible at all. In mukbang, etiquette shifts into recording practice: sounds must be close enough to create intimacy, but controlled enough to remain pleasurable and shareable. Repetition turns these thresholds into expectations, so recording practice begins to function like etiquette.

This point helps explain why mukbang can feel intimate while withholding many of the ordinary conditions of eating together. Shared meals involve sensory and social demands that exceed the sounds of eating. Mukbang reduces those demands while intensifying auditory access to the eating body. The listener is placed close to breath, saliva, texture and throat movement, while remaining distant from the material and social obligations of the shared table. This is a managed form of bodily closeness. It is intimate because the listener is brought near to normally buffered bodily processes. It is mediated because recording choices and platform conventions control what kind of nearness is possible.

This connects mukbang to work on digital intimacy and ambient co-presence, but the consumer culture issue is more specific. Zappavigna (2023) shows how YouTube videos can organise ambient embodied co-presence through visual and aural perspective. Mukbang does something similar around eating. The position of proximity is built through the recorded relation between mouth and food, and through the audible timing of ingestion. Yet the central relation is organised around consumption itself. The format draws the viewer close to another person's eating without requiring shared food, reply, smell, cleanup or reciprocity. Mukbang therefore offers a form of consumer intimacy organised through distance as much as proximity.

This also expands the object of consumption. In mukbang, the viewed meal becomes part of a larger sensory offering. That offering includes the performed breakdown of food, the close recording of mouth work, and the listening position created for the viewer. This helps explain why prior work finds companionship, vicarious pleasure and multisensory appeal in mukbang viewing (Anjani et al., 2020; Song, 2021, 2023). The cases add a more specific account of how such possibilities are sonically organised. Vicarious pleasure in mukbang can therefore include a sensory relation to another person's ingestion alongside imagined taste and visual abundance. Vicarious listening names that relation: another person's consumption becomes followable through the sounds by which it unfolds.

The broader point is that consumer platforms can make bodily processes scalable. Here, scalable means that a private bodily event can be recorded once and encountered repeatedly through replay, search, commentary and sharing. Particular moments, such as a crunch, swallow or breath, can become valued details within the format. Mukbang therefore turns ordinary bodily acoustics into sensory material that can circulate.

This suggests a wider agenda for consumer culture research into how formats make consumer relations audible: how bodies are brought near, how ordinary actions become repeatable scenes for attention, and how formats manage the line between sonic intimacy and discomfort. Future work could examine which bodily or object sounds become desirable, which remain irritating or excessive, and how these boundaries are shaped by platform genre. Mukbang is an especially revealing case because it makes that problem unusually clear. It shows how the ordinary sounds of bodily consumption can be reorganised as a consumer relation that is intimate.

5.4 Scope conditions and limits

The argument is deliberately bounded. It applies most strongly to mukbang content in which ingestion remains close, continuous, and temporally clear. It applies less well to food vlogs organised mainly around talk, highly compressed highlight edits, music-led clips, or recordings where distant capture makes mouth sounds incidental**.** In these cases, other elements may organise the experience more strongly than ingestion sound.

The study is interpretive. It identifies how selected clips format eating as a listenable relation**,** but it does not measure appetite effects, health outcomes, craving reduction, felt fullness, or universal audience responses. It therefore cannot establish that viewers consume through the performer. Claims about proxy consumption, virtual satiation or substitution would require audience-based research rather than close analysis of clips.

These limits do not weaken the argument; they specify its level. It explains the sonic conditions under which accompaniment becomes plausible, while leaving open how different viewers

respond to those conditions. That distinction matters because mukbang is not one format with one effect. It is a varied platform genre in which sound, image, talk, editing and interaction can take different roles. The contribution of this paper is to show what becomes audible when ingestion sound is treated as central to that arrangement.

6. Conclusion

Mukbang makes ordinary eating strange again. Eating belongs to appetite and company, but here it also becomes something to listen through. Ingestion sound is brought forward until it organises a relation between an eating body and viewers elsewhere. The viewer keeps company with the meal's unfolding without sharing the table. This relation is what the paper calls vicarious listening.

The value of the concept is that it names a relation rather than an outcome. It describes how another person's consumption is made available to follow through sound. This keeps the argument close to what the analysis can show while preserving the larger consumer culture question: how do markets organise bodily nearness at a distance?

Patterson and Larsen's sonic turn asks what becomes available when consumption is approached through listening. Mukbang answers that question through the eating body. The meal expands from displayed food into an event arranged through sound. The mouth becomes the listener's point of access, and the microphone turns small bodily transitions into public sensory material. Eating becomes less private, less fleeting and less contained by the table. This matters because mouth sounds have an unstable social status. Manners often push them to the

edges of the meal. Mukbang gives them another setting. Once recorded, framed and repeated, they can become attractive rather than awkward, intimate rather than intrusive, valuable rather than accidental. The body's ordinary acoustics become part of what is consumed.

The wider implication reaches beyond mukbang. A sonic approach to consumer culture begins from the relations that sound makes available: how bodies become near, how ordinary practices acquire rhythm, and how private acts become public enough to circulate. Mukbang is useful because it makes this arrangement unusually exposed. An eating body is recorded, formatted and made available as a way of keeping company. The table becomes a channel. The meal becomes a sequence. Togetherness becomes managed nearness. Mukbang leaves a sharper question for consumer culture research: what kinds of bodily relation do markets make audible enough to consume?

References

Albert, S., & Bell, G. G. (2002). Timing and music. *Academy of Management Review*, *27*(4). https://doi.org/10.5465/AMR.2002.7566078

Anjani, L., Mok, T., Tang, A., Oehlberg, L., & Goh, W. B. (2020). Why do people watch others eat food? An Empirical Study on the Motivations and Practices of Mukbang Viewers. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3313831.3376567

Aucoin, J. (2019). *Virtual Commensality: Mukbang and Food Television.[McGill University, Montreal, Canada]*.

Bradshaw, A., & Holbrook, M. B. (2008). Must we have Muzak wherever we go? A critical consideration of the consumer culture. *Consumption Markets & Culture*, *11*(1). https://doi.org/10.1080/10253860701799959

Chion, M., & Gorbman, C. (1994). The Three Listening Modes. *Audio-Vision - Sound on Screen*.

Choe, H. (2019). Eating together multimodally: Collaborative eating in mukbang, a Korean livestream of eating. *Language in Society*, *48*(2). https://doi.org/10.1017/S0047404518001355

Feld, S. (2015). Acoustemology. *Keywords in Sound*, *12*.

Ferdous, H. S., Ploderer, B., Davis, H., Vetere, F., & O'Hara, K. (2016). Commensality and the social use of technology during family mealtime. *ACM Transactions on Computer-Human Interaction*, *23*(6). https://doi.org/10.1145/2994146

Garlin, F. V., & Owen, K. (2006). Setting the tone with the tune: A meta-analytic review of the effects of background music in retail settings. *Journal of Business Research*, *59*(6). https://doi.org/10.1016/j.jbusres.2006.01.013

Kang, E. K., Lee, J., Kim, K. H., & Yun, Y. H. (2020). The popularity of eating broadcast: Content analysis of "mukbang" YouTube videos, media coverage, and the health impact of "mukbang" on public. *Health Informatics Journal*, *26*(3). https://doi.org/10.1177/1460458220901360

Kavanagh, D. (2013). The limits of visualization. Ocularcentrism and organization. In E. Bell, S. Warren, & J. E. Schroeder (Eds.), *The Routledge Companion to Visual Organization* (1st ed., pp. 1–26). Routledge.

Kytö, M. (2018). Sonic Etiquette: Domestication of Acoustic Neighbourhood Relations in Istanbul. *MSGSÜ Sosyal Bilimler Dergisi*, *2*(17).

Medina, A. M. (2023). The Trans Ear/(h)earing. *Resonance: The Journal of Sound and Culture*, *4*(3). https://doi.org/10.1525/res.2023.4.3.300

Patterson, M., & Larsen, G. (2019). Listening to consumption: Towards a sonic turn in consumer research. *Marketing Theory*, *19*(2). https://doi.org/10.1177/1470593118787583

Poerio, G. L., Blakey, E., Hostler, T. J., & Veltri, T. (2018). More than a feeling: Autonomous sensory meridian response (asmr) is characterized by reliable changes in affect and physiology. *PLoS ONE*, *13*(6). https://doi.org/10.1371/journal.pone.0196645

Reinoso Carvalho, F., Wang, Q. (Janice), van Ee, R., Persoone, D., & Spence, C. (2017). "Smooth operator": Music modulates the perceived creaminess, sweetness, and bitterness of chocolate. *Appetite*, *108*. https://doi.org/10.1016/j.appet.2016.10.026

Ryan, G., & Ruiz, M. (2026). The sound of consumption: The electromagnetic soundwalk as anti-method. *Organised Sound*. https://doi.org/10.1017/S1355771825100988

Sanskriti, S., Guglani, I., Joshi, S., & Anjankar, A. (2023). The Spectrum of Motivations Behind Watching Mukbang Videos and Its Health Effects on Its Viewers: A Review. *Cureus*. https://doi.org/10.7759/cureus.44392

Smith, S. (2025). Running thoughts through music. A musical inflection of the sonic turn in consumer research. *Marketing Theory*. https://doi.org/10.1177/14705931251396383

Song, H. G. (2021). *A study on Mukbang user's consuming behavior*. University of Nevada, Las Vegas.

Song, H. G. (2023). Understanding Social Media Users' Mukbang Content Watching: Integrating TAM and ECM. *Sustainability (Switzerland)*, *15*(5). https://doi.org/10.3390/su15054013

Spence, C. (2015). Eating with our ears: assessing the importance of the sounds of consumption on our perception and enjoyment of multisensory flavour experiences. *Flavour*, *4*(1). https://doi.org/10.1186/2044-7248-4-3

Spence, C., & Wang, Q. (2015). Sensory expectations elicited by the sounds of opening the packaging and pouring a beverage. *Flavour*, *4*(1). https://doi.org/10.1186/s13411-015-0044-y

Sterne, J. (2012). *MP3 The Meaning of a Format*. Duke University Press.

Turley, L. W., & Milliman, R. E. (2000). Atmospheric effects on shopping behavior: A review of the experimental evidence. *Journal of Business Research*, *49*(2). https://doi.org/10.1016/S0148-2963(99)00010-7

Wang, C., Peng, Y., Qiu, L., & Wan, X. (2021). Cloud-Based Commensality: Enjoy the Company of Co-diners Without Social Facilitation of Eating. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.758966

Zampini, M., & Spence, C. (2004). The role of auditory cues in modulating the perceived crispness and staleness of potato chips. *Journal of Sensory Studies*, *19*(5), 347–363.

Zappavigna, M. (2023). Digital intimacy and ambient embodied copresence in YouTube videos: construing visual and aural perspective in ASMR role play videos. *Visual Communication*, *22*(2). https://doi.org/10.1177/1470357220928102