


ARTICLE

# ‘This you may <sub>NNNN</sub>NEVER have heard before’: initial lengthening of accented negative items as vocal-entangled gestures

Simon Harrison 

Department of English, City University of Hong Kong, Kowloon, Hong Kong  
Email: [sharriso@cityu.edu.hk](mailto:sharriso@cityu.edu.hk)

(Received 28 January 2023; Revised 01 April 2024; Accepted 04 April 2024)

## Abstract

Movement scientists have proposed to ground the relation between prosody and gesture in ‘vocal-entangled gestures’, defined as biomechanical linkages between upper limb movement and the respiratory–vocal system. Focusing on spoken language negation, this article identifies an acoustic profile with which gesture is plausibly entangled, specifically linking the articulatory behaviour of onset consonant lengthening with forelimb gesture preparation and facial deformation. This phenomenon was discovered in a video corpus of accented negative utterances from English-language televised dialogues. Eight target examples were selected and examined using visualization software to analyse the correspondence of gesture phase structures (preparation, stroke, holds) with the negation word’s acoustic signal (duration, pitch and intensity). The results show that as syllable–onset consonant lengthens (voiced alveolar /n/ = 300 ms on average) with pitch and intensity increasing (e.g. ‘<sub>NNNN</sub>NEVER’), the speaker’s humerus is rotating with palm pronating/adducting while his or her face is distorting. Different facial distortions, furthermore, were found to be entangled with different post-onset phonetic profiles (e.g. vowel rounding). These findings illustrate whole-bodily dynamics and multiscalarity as key theoretical proposals within ecological and enactive approaches to language. Bringing multimodal and entangled treatments of utterances into conversation has important implications for gesture studies.

**Keywords:** biomechanics; entanglement; facial expression; gesture; languaging; lengthening; multimodality; multiscalarity; negation; stress

## 1 Introduction

This article responds to the call for ‘simultaneous assessment of prosodic and gestural components of language’ (Esteve-Gibert et al., [this issue](#)) with a study of the material bodily dynamics that give rise to the experience and contours of dialogic spoken

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



language communication, taking for a case study the morphosyntactic expression and distinctive gestures of negation.

The linguistics and pragmatics of negation have proven to be a fertile ground for identifying the prosodic and gestural components of language and assessing their interrelations. What Prieto and Espinal (2020) call ‘multimodal negation’ refers to ‘the integration between prosody and gesture in the expression of negation’ (p. 690), noting in their research synthesis that ‘prosodic and gestural patterns may be used in combination with negative expressions (*no, not, nothing*, etc.) to express denial, rejection, and related notions like correction, disapproval, stop, prohibition, failure, refusal, etc.’ (p. 690; see also Boutet et al., 2021, p. 32). There are distinct strands of research dedicated to spoken language negation from prosodic perspectives (e.g. Blanchette and Nadeau, 2018; Kaufman, 2002) and gestural perspectives (e.g. Bressemer & Müller, 2017; Calbris, 2011; Harrison, 2018; Kendon, 2004; Müller, 2017). Multimodal negation along the lines of Prieto and Espinal (2020) is currently being explored with various methods in different domains. These include psycholinguistic experiments evaluating how people interpret negative sentences when presented with multimodal stimuli that manipulate the coordination of linguistic forms, gestures and prosodic contours (Li et al., 2016; Prieto et al., 2013; Prieto & Espinal, 2020; Tubau et al., 2015). Another is video-based studies of multimodal negation in child–parent interactions using micro-analysis and corpus methods to track language development processes during real-time spoken and signed language (Beupoil-Hourdel, 2015; Beupoil-Hourdel & Morgenstern, 2021; Boutet et al., 2021).

The tradition of multimodality evoked here grows out of dissatisfaction with models of language and communication that have historically focused only on speech. That ‘Human interaction does not occur only on the verbal dimension’ is by now an oft-heard critique in the opening statements to studies of speech and gesture (Brown & Prieto, 2021, p. 430). I agree with (Rohrer et al., 2023) that contemporary scholars of language and communication are much more likely to ‘acknowledge that language is a multimodal phenomenon where multiple modes of communication (e.g., auditory and visual modes) come together to express meaning’ (p. 7). I also think that referring to language as ‘multimodal’, to hearing and vision as ‘modes’ coming together and to meaning as being ‘expressed’ (from Latin to *press* or *push out*) creates an object of study and form of enquiry that we can similarly acknowledge and question. One aim of this article is to integrate different modes of inquiry and research traditions that open new perspectives on prosody and gesture for phenomena commonly referred to as ‘multimodal’.

In her monograph *Elements of Meaning in Gesture*, French semiotician Calbris elaborates on this expanded view of communication as multimodal as follows:

One uses parallel sensory pathways, audio-oral and visual-gestural, which interact in multimodal communication, that is, the ensemble of spoken linguistic, prosodic, intonational, gestural, postural, and facial activity that participants engage in when they ‘talk’. The spoken linguistic, prosodic, and intonational activity employs the audio-oral modality; and the gestural, postural, and facial activity employs the visual-gestural modality (2011, pp. 5–6).

This ensemble of activity involved in ‘talk’ (understood as a form of social organization) is what gesture studies pioneer Kendon defined as the speaker’s ‘utterance’.

'In creating an utterance that uses both modes of expression', Kendon (2004) concluded from fine-grained analyses of speech and gesture in social interaction, 'the speaker creates an ensemble in which gesture and speech are employed together as partners in a single rhetorical enterprise' (p. 127). This multimodal ensemble view of the relation between speech and gesture – utterances as 'units of activity (that) may be constructed from speech or from visible bodily action or from combinations of these two modalities' (Kendon, 2004, p. 7) – has proven influential in structuring the empirical observation and modelling of language and communication. In the domain of negation, for instance, these communicative–semiotic approaches to multimodality have led researchers to foreground the identification of linguistic items, prosodic contours, facial expressions and gesture forms or flows that participate in the situated expression of negation and its embodied understandings. Multimodal negative utterances have accordingly been conceptualized as combinations or ensembles of such components (indeed 'units' or 'constructions' from a cognitive–linguistic perspective; Bressemer & Müller, 2017; Cienki, 2012; Harrison, 2018) that speakers orchestrate or integrate to conventionally encode and express various (negative) functions, meanings and feelings during embodied language use and situated interaction. My previous analyses of vertical palm gestures with negative utterances involving notions of blockage, force and distance exemplify this position, leading me to conclude that 'a multimodal utterance is the expressed conceptualisation in structured and content-laden form, composed of both verbal and gestural components' (Harrison, 2018, p. 77). However, research from other fields and disciplines helps to see that communication and semiosis (along with encoding and expression) are only one level or scale of prosody and gesture's integration.

Consider that Pouw and colleagues' (Pouw & Fuchs, 2022) 'gesture–speech biomechanics thesis' stems from empirical evidence that '*when an upper-limb segment with a certain mass (or multiple segments with a certain combined mass) sufficiently accelerates or decelerates, it yields physical impulses on the musculoskeletal system, the cascading mechanical effects of which will constrain respiratory–vocal activity*' (p. 4; emphasis original).<sup>1</sup> As experiments reviewed by Pouw et al. (2002a) have shown, these constraints contribute to many of the acoustic effects traditionally measured and perceived for prosody, including change in duration, fundamental frequency and intensity of phonation. The thesis is that 'gesture–speech synchrony may be grounded in biomechanical linkages between upper limb movement and the respiratory system' (Pouw, Harrison, et al., 2020b, p. 1243). The researchers refer to these linkages as 'vocal-entangled gestures' (Pouw & Fuchs, 2022), arguing that 'such gestures have communicative potential by serving as an index for some embodied state of affairs rather than as a representation of purely mental content' (p. 1).

If prosody and gesture are configured biomechanically, their entanglement should open new scales and dynamics for studying the kinds of utterances characterized as multimodal negation. Likewise, the phenomenon known as multimodal negation should provide an ecological context where attested biomechanical processes can be inferred and where new vocal-entangled gestures can potentially be identified for future biomechanics research. Yet multimodality and entanglement are different ways of approaching and representing the world. Their concepts rely on

<sup>1</sup>I am grateful to gesture philosopher Cuffari whose work first drew my attention to biomechanical research (Cuffari, 2022).

different forms of scientific analysis and ways of thinking, leading scholars to different definitions of utterance and notions of social understanding.

For researchers of multimodality unfamiliar with ‘entanglement’, the term might evoke ‘some kind of mess/random organization’ at odds with the systematic combining and orchestrating of components into ensembles and constructions familiar from multimodality research.<sup>2</sup> However, entanglement is a term that offers an empirical and theoretical perspective on relations between dynamical systems (e.g. Hepp and Couldry, 2023; Pendleton-Jullian & Brown, 2023; Scott and Orlikowski, 2014), which include ourselves and each other. In philosophy of cognitive science, for instance, when Di Paolo et al. (2018) introduce this term to discuss ‘patterns of social interaction’, they say that people in co-presence become ‘entangled’, by which they refer to ‘the presence of deep correlations between processes at multiple timescales in each body’ that range from ‘[i]nter-brain synchronisation at high frequencies’ to much slower ‘emergent rhythms of social interaction (e.g., the movements of the hands, the taking of turns, and so on)’ (p. 77). To think of interaction as arising from processes at different scales and speeds happening in and throughout co-present bodies, rather than (or in addition to) arising from the coming together of modes with which people express their meaning, offers one way to distinguish entangled and multimodal treatments of gesture.

In studying how ‘the speaker creates an ensemble in which gesture and speech are employed together’ (Kendon, 2004, p. 127), for instance, Kendon’s landmark analyses work on a scale and speed in which ‘the gesture and speech components of an utterance’ (p. 156) can be clearly distinguished, classified and attributed to an autonomous agent – as required by the transcription practices that developed along with this approach (pp. 362–364). The research into ‘vocal-entangled gesture’, on the other hand, reveals correlations at scales and speeds that are happening ‘deeper’ and/or at ‘higher frequency’ and which have their own vitality and animacy. Such correlations include those within the bio-anatomy of our respiratory–motor systems discovered experimentally with various muscle, posture, motion and audio-sensing technologies (Pouw et al., 2024; Pouw & Fuchs, 2022). When compared to the neat partitioning of components or activities into audio–oral and visual–gestural modes from where multimodal treatments of speech and gesture often set out (cf. Calbris, 2011), these entangled correlations (of fleshy tissues, fibres, liquid and breath) will seem ‘messy’. But their organization as we speak and gesture is certainly not ‘random’, instead being described as non-linear, self-organizing, fractal, cascading and dynamic (Di Paolo et al., 2018; Thibault, 2021). Utterances from this perspective, therefore, are less of an ensemble of distinct components, activities or resources orchestrated by the speaker to encode and express thoughts and feelings. ‘Utterances’, to quote languaging scholar Thibault (2021), ‘are best seen as synergies of bodily movement that we learn to harness in order to move others and to be moved ourselves by others...; activities arising out of the interactions between sub-processes of organism-persons’ (p. 194). The interactions between the forelimb motion and respiratory–vocal systems identified by Pouw and Fuchs as ‘vocal-entangled gestures’ would exemplify such synergies of movement. These processes of entanglement force us to keep in mind that utterances are

<sup>2</sup>I thank an anonymous reviewer for raising this concern.

‘activities of living, animated, feeling organism-persons’ (Thibault, 2021, p. 195), to keep in mind that *utterances are us*.<sup>3</sup> From this perspective, the ‘cascading mechanical effects’ from people’s gestures onto their bodies’ musculoskeletal and vocal–respiratory systems must also be acknowledged as *felt* affects, not only biomechanical but also bio-*kinesthetic*. Another aim of this paper is to further conceptualize and illustrate how multimodal and entangled treatments of prosody and gesture can relate and inform each other.

With examples from a video corpus featuring media personalities, intellectuals and comedians, this research explored negative utterances as vocal-entangled gestures and understood here as the correspondence of material bodily dynamics that give rise to or announce themselves as a particular grammatical–gestural pattern.<sup>4</sup> The approach to language and cognition developed here thus aims to de-emphasize the view of speech and gesture as components within modes that ‘come together’ into utterances as ensembles and to focus instead on the ‘dynamical field of bodily energy’ or ‘synergies of bodily movement’ from which spoken utterances grow forth and materialize (Thibault, 2021, p. 194). The article works to shift perspective and scale from the forms and organization properties commonly identified for multimodal ensembles to the immanence and bodily feeling of material dynamics at the ‘pico-scale’ of first-order languaging activity. My study of negation illustrates how these two scales and their ways of working may interrelate.

The term ‘pico-scale’ is borrowed from approaches to language and cognition known as ‘bio-social’, ‘ecological’ and ‘distributed’ (see Thibault, 2011, 2021). Here, researchers are developing distinctions not between different modes but between different *orders* of language that enable and shape communicative situations, namely between ‘first-order languaging’ and ‘second-order language’. As introduced in an important paper in *Ecological Psychology*, first-order languaging refers to ‘synchronized interindividual (whole-)bodily dynamics on very short, rapid timescales of the order of fractions of seconds to milliseconds’, while second-order language refers to more ‘stabilized cultural patterns on longer, slower cultural timescales’ such as ‘lexico-grammatical patterns’ (Thibault, 2011, pp. 214–216). In Thibault’s (2021) relational ontology of languaging, the first-order languaging activities include pico-scale dynamics that ‘are constantly emerging and coalescing with other features on larger scales so as to give rise to the familiar dynamical properties of utterances’ (p. 183). One impetus of languaging research is to shift focus to first-order activities in order to redress the ‘written language bias’ that has dominated linguistics (Linell, 2011/2005) and to ‘open up new ways of thinking about the organisation and functionality of languaging’ (Thibault, 2021, p. 12). Another goal is to develop an explanation for how first-order (pico-scale) languaging and second-order (cultural–historical) language relate. By exploring the material bodily dynamics that give rise to the experience and contours of grammar–gesture patterns recognized as multimodal negation, this paper’s findings can contribute to accounts of ecological languaging and multiscalearity.

<sup>3</sup>I detour this expression from affect theorist Massumi (2015), who writes of our bodily experiences: ‘They’re us, they’re what we’re made of. We are our situations, we are our moving through them. We are our participation – not some abstract entity that is somehow outside looking in at it all’ (pp. 13–14).

<sup>4</sup>My preference for describing prosody–gesture relations as a ‘correspondence with’ rather than an ‘interaction between’ is inspired by the phenomenology of Ingold (Ingold, 2021).

More specifically, my analyses of negative utterances using different visualization technologies (European Distributed Corpora Project Linguistic Annotator (ELAN) and Phonetic and Acoustic Analysis Toolkit (PRAAT)) led me to discover salient pico-scale dynamics. As I will show, prior to the accented negative items known to synchronize with the stroke of gestures associated with negation and align with negation's grammatical constraints (Harrison, 2010; 2014a; 2018), a range of acoustic and kinesthetic phenomena visible in the spectrogram and inferred from the video feed are *already corresponding*. In the examples to be presented in this article, as the speaker's forearm undergoes internal–external rotation of the humerus (a 'lateral sweep' gesture is being prepared), there is lengthening of the negative word's accented onset (the voiced alveolar nasal consonant /n/). During this time, there is also increased vocal fold vibration and pulmonic flow (rising F0 and intensity), as well as distortion of facial muscles. In other words, while the forelimb gesture prepares for its stroke we also see changes in the vocal acoustics that materialize in the pitch and intensity peaks of the emphasized negation particle ('NNNNNEVER'). Changes in the superficial musculoaponeurotic system involved with facial expressions are also initiated here. Rather than conceptualizing this activity as the speaker creating utterances by coordinating, combining or otherwise orchestrating these elements into an ensemble, in Thibault's (2021) take on languaging, what we see here are the whole-bodily pico-scale material dynamics emerging and coalescing for (and being constrained 'top-down' by) the lexico-grammatical pattern and culturally distinctive ('recurrent') gestures of negation.<sup>5</sup> Furthermore, since my examples all involve gestures whose kinesiology and cascading mechanical effects on respiratory-related muscle systems have been discussed in the biomechanics literature ('lateral sweep'), we can infer that the specific correspondence of prosody and gesture described in this paper plausibly involves vocal-entangled gestures. We can likewise reasonably speculate that the superficial musculoaponeurotic system pulling facial expressions is also entangled with limb movement and the respiratory system. Recent research demonstrates that 'the voice interacts with muscle chains of *the whole body*' (Pouw et al., 2024, p. 10; emphasis added).

The correspondence of the articulatory behaviour of onset–consonant lengthening with forelimb gesture preparation and facial deformation discovered in the present study has not yet been observed in multimodal negation or gesture–speech biomechanics research, yet comes into view by bringing multimodal and entangled treatments of utterances into conversation. While contributing empirically to prosodic research on lengthening and gesture research on negation, the interplay of multimodality and entanglement developed in this paper illustrates whole-bodily dynamics and multiscalarly as key theoretical proposals of ecological and enactive approaches to language (Di Paolo et al., 2018; Harrison, in preparation; Thibault, 2021). The findings also pave the way for future biomechanical and psycholinguistic testing.

The rest of the article is structured by the background on multimodal negation, from which I build a bridge to research on speech–gesture biomechanics, followed by a methods section explaining how examples were collected and analysed. I then present the analyses and finally discuss the findings.

<sup>5</sup>On 'recurrent gestures', see Harrison & Ladewig (2021).

## 2 Multimodal negation

As a universal linguistic context, negation has historically provided an excellent point of departure for a systematic exploration of prosody and gesture from a multimodal perspective (see, e.g., Prieto & Espinal, 2020). A brief introduction to negation from a traditional linguistics perspective as well as from prosodic and gestural perspectives will be useful before discussing studies that have explored how linguistic, prosodic and gestural patterns of negation relate. I then make the link to biomechanics research that justifies exploring accented negative items as vocal-entangled gesture.

### 2.1 Linguistics of negation

Horn (1989) cites ‘the traditional criteria for negativity’ in linguistic communication for some linguists as being ‘the presence of a negative particle, its appearance in a specified syntactic location, and so forth’ (p. 34). Negation in English, for example, is ‘marked by individual words (such as *no*, *not*, *never*) or by affixes within a word (such as *-n’t*, *un-*, *non-*)’ (Huddleston & Pullum, 2005, p. 149; original emphasis). Syntactically, these negative items tend to occur as early in the sentence as possible and to be reinforced (Jespersen, 1917), leading to related linguistic phenomena like negative polarity items, such as *any*, *even*, *ever* and *at all* (Lawler, 2005). Introducing negative forms and their contractions to an utterance is known to impose positional constraints on other elements ‘downstream’ in syntax through constructs and operations that include the ‘scope’ and ‘focus’ of negation (Horn, 1989). The scope of negation refers to the spread of negation through the utterance following a negative particle or item and indicates as per Huddleston and Pullum (2005) ‘the part of the sentence that the negative applies to semantically’ (p. 156), while the focus of negation denotes ‘the part of that scope that is most prominently or explicitly negated’ (Huddleston and Pullum, 2002, p. 790). For Horn and Wansing (2022), ‘Negation is in the first place a phenomenon of semantic opposition’ (p. 1). It can be classified from a range of perspectives, such as grammatical–typological (e.g. verbal/non-verbal, clausal/sub-clausal, analytic/synthetic, ordinary/meta-linguistic; Huddleston and Pullum, 2002; Miestamo, 2017) and pragmatic, which relates to the functions of negation including ‘denial, rejection, and related notions like correction, disapproval, stop, prohibition, failure, refusal, etc.’ (Prieto & Espinal, 2020, p. 667; see also Beaupoil-Hourdel et al., 2016, pp. 98–99). A number of these aspects of linguistic negation have appeared in studies of prosody.

### 2.2 Prosody and negation

Prosody is understood by Brown and Prieto (2021) as ‘vocal effects that accompany the sounds of individual segments of speech, and that extend over words, phrases or utterances’ noting how prosody ‘allows for the same word, phrase or utterance to be delivered in different ways, such as louder/quieter, faster/slower, with higher or lower pitch, or with different intonation contours’ (pp. 431–432). Such vocal effects arise from changes in the speaker’s body and are physically manifest in the acoustic signal (sound wave) of speech. What we perceive as pitch and volume, for instance, refers to characteristics of this sound wave affected by different articulatory and physiological efforts as well as by the environment or medium through which sound travels. Pitch height often refers to changes in the frequency of the sound wave due to the rate at

which the speaker's vocal folds are vibrating in the larynx known as fundamental frequency (F0), measured in the acoustic signal as hertz (Hz) or cycles per second. This cycle of vibration involves the opening and closing of the vocal folds in the glottis (the 'glottic cycle'). The quicker these vocal folds vibrate, the higher the perceived pitch, though factors other than F0 may contribute to the perception of pitch. Changes in volume (more technically 'loudness') refer to changes in the amplitude or height of the sound wave, which represents the intensity of acoustic energy measured in decibels (dB). Intensity can be affected by the relative amount of air coming through the glottis from the speaker's lungs (pulmonic flow) and his or her 'vocal effort' or 'force' (i.e. shouting).

In their research overview, Prieto and Espinal (2020) synthesize studies of the relation between prosody and negation markers as well as negative node and scope. They discuss studies of interactive speech showing that 'the locus of negation is almost always marked with a high pitch accent' (p. 681) and how 'prosodic patterns strongly interact with sentence interpretation' (p. 684). The tendency of negative particles to be accented syllables brings a range of prosodic phenomena into view such as *stress*, i.e. the 'property of a syllable which may, depending on the language, mark it as more prominent than its neighbours' (De Jong, 1995, p. 491). According to De Jong's (1995) overview of linguistic stress, syllables that are stressed arise from a range of articulatory behaviour in the speaker's vocal tract: articulatory gestures are more flexible (as opposed to stiff), sharper (exhibits less coarticulation effects), louder (increased intensity) and longer (aimed at increased sonority). This accentual lengthening is reviewed by Turk and White (1999), who note that while 'an F0 excursion associated with a lexically stressed syllable is considered to be the primary cue to phrasal stress, duration is considered to be an important secondary cue' and, furthermore, that 'initial consonants showed a greater degree of lengthening than final consonants' (p. 173). For Klatt (1976), segmental duration was not secondary but 'one of the primary cues to the existence and location of emphasized or contrastively stressed material' (p. 1219).

Lengthening of syllable onset consonants for emphasis (as opposed to for contrast) is explored in Niebuhr's (2010) study of intensifying emphasis in German. Through studies of read speech with 'target' syllables indicated for stress, lengthening was linked to a form of semantic-pragmatic emphasis called 'negative intensification' with statistically reliable effects. Greater lengthening of consonants preceded short vowels compared to long ones, for example. 'Emphasis for intensity means the semantics of the word that bears the emphasis is intensified', with negatively valenced intensity occurring with words such as 'stupid', 'drunk', 'idiotic' and 'lousy' (pp. 172–173).<sup>6</sup> The documented phonetic profiles of accented syllables were the basis for Niebuhr to propose a form–meaning correspondence of consonant lengthening and negatively valenced sentiment, which Ward (2019) subsequently relates to iconicity in his typology of 'likely primordial mappings' of prosodic meaning (pp. 57–60). However, the target syllables in that study did not include explicit negative particles, so the intuitive link between negative intensification and linguistic negation was missed

---

<sup>6</sup>Thibault (2021) notes that valence is a dimension of emotion defined as its 'positive or negative attractiveness' (p.189) and presents analyses arguing that prosodic contours reveal 'the establishment and subsequent working through of an affective valence' as speakers' bodies negotiate feelings and meanings (p. 186).



(words such as ‘stupid’, ‘drunk’ and ‘idiotic’ have been typologically classed as harbouring ‘implicit’ or ‘covert’ negation).

Although articulatory behaviours such as lengthening have long been referred to as ‘gestural’ (cf. ‘gestural theory of speech production’; De Jong, 1995) and can be induced in ‘virtually the same’ way by the occurrence of rhythmic manual, head or eyebrow ‘beat’ gestures with implications for perceived prominence (c.f., Kraemer & Swerts, 2007, p. 42), they have not been closely related to the study of other bodily gestures associated with negation.

### 2.3 Gesture and negation

Rich observations on gesture’s links with negation have long featured in the work of key thinkers and texts on human communication, such as in Darwin’s (1872) *The Expression of Emotions in Man and Other Animals* and in the *Institutio Oratoria of Quintilianus* (see Kendon, 2004). Examples of gestures that have been associated with spoken language negation include the headshake (with its notorious cultural variations; Harrison, 2014b) and facial expressions like the ‘not face’ (Benitez-Quiroz et al., 2016). However, various manual gestures striking outwards from the body or raised with open palms towards the addressee have received much more attention and analysis in gesture studies (Boutet et al., 2021; Bressemer & Müller, 2014; 2017; Calbris, 2005; 2011; Harrison, 2018; Kendon, 2004; Lapaire, 2006).

Multimodal treatments of these gestures have demonstrated how they interrelate and work together with the linguistics and pragmatics of negation during spoken communication (for overviews, see Harrison, 2024; Prieto & Espinal, 2020). Quite consistently across several languages, gestures have been kinesically related to the dialogic contexts of use in which negation is being expressed (Kendon, 2004) as well as with the conceptual semiotics and cognitive etymology of negation (Calbris, 2011; Lapaire, 2006). Building on these authors and working with a corpus of German speech, Bressemer and Müller (2014) identified different patterns of gestures whose forms can be interpreted as cognitively and semiotically motivated by underlying actions that all involve movements away from one’s body (sweeping away, holding away, brushing away and throwing away) and showed their association with different negative meanings and functions. Calbris (2011), on the other hand, treats gestures like ‘sweeping away’ and ‘holding away’ as examples of ‘polysemous gestures’ with ‘plural motivation’. In the case of ‘sweeping away’, its meaning depends on the component of the level hand that is salient or profiled by the given context (such as its movement trajectory or shape configuration). This determines the analogical link established with a physical correlate, which in turn is the source of the gesture’s meaning and array of subsequent semantic derivations (Calbris, 2011, p. 183). Based on a purely physiological analysis of this gesture (‘lateral sweep’), however, Boutet (2015, 2018) argues that the source of its core meaning is generated on the hand.

Focusing on the timing and organization of such gestures with respect to the grammatical structures of co-occurring negative sentences, Harrison (2018) worked out a heuristics for the synchronization of gesture with respect to linguistic negation, presenting evidence that the positional constraints imposed by grammatical negation, namely negative node, scope and focus, also extend to the organization and timing of gesture (Chapter 3). Although negative particles involve accented syllables (cf. Section 2.2), which Rohrer et al. (2023) include among ‘prosodic landmarks for

gesture production' (p. 26), these heuristics did not link the temporal constraints on gestures associated with negation to prosody. In fact, the relationship between prosody and gesture was not considered in-depth by the above gesture studies of negation, including by researchers proposing the universal 'not face' (Benitez-Quiroz et al., 2016).

#### 2.4 Multimodal negation

The interplay of grammatical, gestural and prosodic patterns of negation has been specifically explored in at least two lines of studies. Focusing on the comprehension of negative sentences that cognitively are not straightforward to process, a programme of psycholinguistic research has explored how the coordination of prosodic and gestural patterns mentioned above may affect the interpretation of negation and their role in ambiguity resolution (Brown & Kamiya, 2019; Li et al., 2016; Prieto et al., 2013; Prieto & Espinal, 2020; Tubau et al., 2015). Influenced by previous studies of gestural negation, researchers draw their patterns of linguistic, gestural and prosodic forms from elicited utterances and then create audio-visual stimuli that manipulate the coordination of different resources. Through a perceptual study of sentences that are ambiguous as to single or double negation, for example, Prieto et al. (2013) demonstrated that 'prosodic and non-verbal cues (i.e. gestural patterns) crucially affect the interpretation of isolated n-words' (p. 147). This effect has been explored in the interpretation of answers to negative yes/no questions in Catalan (Tubau et al., 2015), of rejections to negative assertions/questions in Mandarin Chinese (Li et al., 2016) and of negation and quantification (Brown & Kamiya, 2019). In these studies, negation is treated as a linguistic meaning or function and its understanding is conceived of as the perceiver's ability to make the correct inference or judgement of the speaker's intended meaning based on integrating multisensory stimuli, although not taking into account bodily feeling.

An example from Beaupoil-Hourdel and Morgenstern's (2021) recent study of 'shrug' gestures in family interactions can illustrate the findings from a line of multimodal negation research during language development. An exasperated child responds to persistent questioning from Mum concerning the whereabouts of a toy with 'euh je sais pas!' (*er I don't know!*; lit. I know not). She 'couples' this utterance with a 'composite gesture' comprising 'three distinct forms: a palm-up on both hands, a shoulder lift, and a head tilt' (p. 212). Analysing this gesture kinesiolgically (Boutet, 2015, 2018; Morgenstern et al., 2021) 'as movement that flows from one body part (a segment) to the next' (p. 188), and observing this flow's relation with prosody and lexico-grammar, Beaupoil-Hourdel and Morgenstern explain that the part 'je sais' (*I know*) 'follows a rising prosodic contour and is produced timely with the scope of the shoulder lift', while the negation word "pas" follows a falling contour with the lengthening of the vowel when Madeleine's body collapses' (p. 212). The researchers analyse the gesture's significance in terms of its function, which they interpret by 'taking into consideration the context of interaction, the immediate previous utterance and the feedback and recast provided by the co-speaker', attributing a designated meaning according to their coding scheme as 'absence', 'affective' and/or 'epistemic' (Beaupoil-Hourdel & Morgenstern, 2021, pp. 189–190). Coupling of prosody and gesture in such utterances is viewed as evidence of the child's increasing control of distinct symbolic resources and their integration along a social,

cultural and cognitive developmental trajectory towards sophisticated multimodal utterances ('the blossoming of multimodal negation'; Beaugoil-Hourdel et al., 2016, p. 99).

We can now turn to research that views the coupling of prosody and gesture from a different perspective, that of 'interactions between sub-processes of organism-persons' (Thibault, 2021, p. 194), namely between forelimb motion and respiratory–vocal systems.

### 2.5 Gesture–speech biomechanics and vocal-entangled gestures associated with negation

Research into gesture–speech physics has conducted experiments and synthesized findings from neural processes, biomechanics and social interaction to develop a 'multilevel multimodal approach' to gesture and prosody (Pouw et al., 2021; cf. Pouw, de Jonge-Hoekstra, et al., 2020a; Pouw, Harrison, et al., 2020b, Pouw & Fuchs, 2022, Pouw et al., 2024). Of interest here is the 'gesture–speech biomechanics thesis', which stems from empirical evidence that 'when an upper-limb segment with a certain mass (or multiple segments with a certain combined mass) sufficiently accelerates or decelerates, it yields physical impulses on the musculoskeletal system, the cascading mechanical effects of which will constrain respiratory–vocal activity' (Pouw & Fuchs, 2022, p. 4). The thesis is that 'gesture–speech synchrony may be grounded in biomechanical linkages between upper limb movement and the respiratory system' (Pouw, Harrison, et al., 2020b, p. 1243). The researchers refer to these linkages as 'vocal-entangled gestures', showing them to be 'present in ontogeny, have deep roots in phylogeny, and have natural communicative significance' (Pouw & Fuchs, 2022, p. 13).

As gesture–speech physics and related experiments have shown, these constraints contribute to many of the acoustic effects traditionally measured for prosody, even unintentionally. These include change in fundamental frequency, intensity of phonation and duration. In one oft-cited study, for instance, Krahmer and Swerts (2007) asked participants to vary how they say 'Amanda gaat naar Malta' (*Amanda goes to Malta*) by placing 'acoustic pitch accent' and/or performing a 'visual beat gesture' on either 'Amanda' or 'Malta'. Their analyses of these target words in PRAAT show that pitch accents and beat gesture led to 'virtually the same' acoustic effects, such as increased duration of the /a/ segments in the syllable receiving accent or beat (amAnda/mAlta). The mechanical effects found to cascade from gestural forelimb movements affect the musculoskeletal system in other ways too, yielding a range of peripheral actions called anticipatory postural adjustments. For example, 'performing an upper limb movement recruits a whole kinetic chain of muscle activity around the trunk (e.g., the rectus abdominis) to maintain posture' (Pouw, de Jonge-Hoekstra, et al., 2020a, p. 91). In experiments where subjects were fitted with respiratory belts, gesturing was also found to affect the respiratory system (Pouw, Harrison, et al., 2020b). The lungs and the ribcage are 'so tightly connected' and separated only by a 'fluid-filled pleural space, which acts as a vacuum', so 'any motion affecting the ribcage, like moving the arm during gesturing, can have a (small) effect on lung volume and subglottal pressure' (Pouw & Fuchs, 2022, p. 4; emphasis original). The fact that 'actions happening locally' like gesture 'can reverberate more globally' is referred to as 'tensegrity' (Pouw & Fuchs, 2022, p. 4). Multiple factors influence how any single gesture reverberates through the body. Those demonstrated experimentally include

size of gesture ('higher-impulse arm movement vs lower impulse wrist movement'), deceleration/acceleration of gesture and posture of speaker ('gestural effects on acoustics were more pronounced when participants were standing vs sitting'; Pouw, Harrison, et al., 2020b, p. 1232).

Happily, Pouw and Fuchs (2022) use the negative utterance 'no way' to illustrate the cascading mechanical effects on respiratory-related muscle systems of a gesture described anatomically as 'external-internal rotation of the humerus' (p. 3; emphasis original). In their illustration (cropped and reproduced in Figure 1), this utterance involves a visibly stressed negative item ('no') and a gesture recognizable from studies of multimodal negation as a 'lateral sweep' (cf. Section 2.3). The physical impulse (PI) of this gesture is visualized by emphasis arrows at the end-point of the gesture's sweep, which corresponds with the 'apex' of a unidirectional gesture stroke (Rohrer et al., 2023). The dashes annotated underneath 'no way' suggest that the full gesture excursion (i.e. the 'gesture unit'; Kendon, 2004) was organized with respect to the lexico-grammar of negation (a 'grammar-gesture nexus'; Harrison, 2018; Lapaire, 2011). The image is montaged alongside the 'myofascial chain' of connective tissues running through the different muscle groups potentially recruited by this gesture.

Circling back to multimodal studies of negation, relevant here is the kinesiological perspective on gesture developed by Boutet (2018), who has analysed the articulatory (joint) physiology of this gesture form and its association with negation (Boutet, 2015; Boutet et al., 2021). In addition to the external-internal rotation of the humerus (upper arm bone) indicated by Pouw and Fuchs, Boutet (2015) specifies that such gestures 'present a position or movement of pronation and adduction of the hand' (p. 126). Boutet's analyses further describe articulatory reverberation in arguing that 'the movement propagation flow [along the upper limb] seems to be distal-proximal' (p. 126), i.e. from hand to shoulder. This is

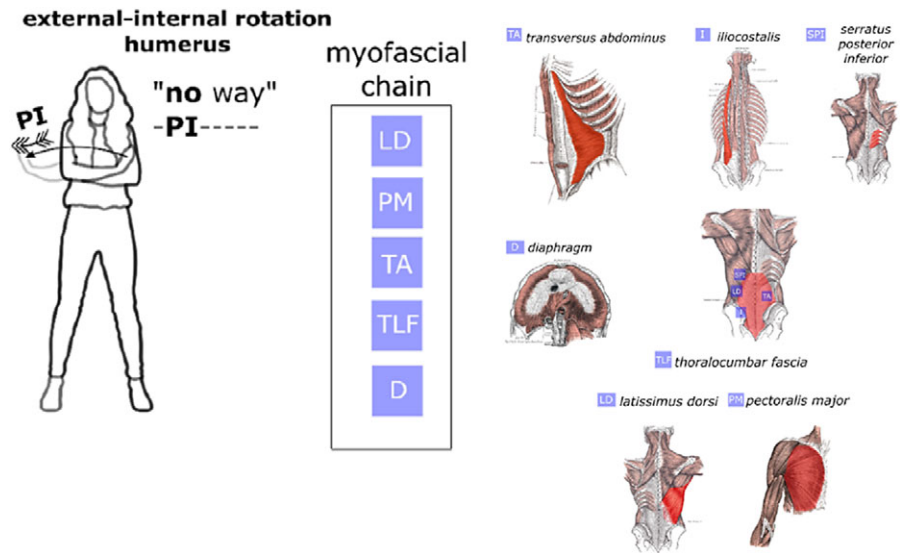


Figure 1. The biomechanics of a well-known multimodal negation (images from Pouw & Fuchs, 2022, p. 3).

Boutet's basis to argue that this gesture's attested negative meanings (rejection, refusal, negation) 'emanate' from the hand.

While Boutet's kinesiological approach associates articulatory physiology with felt-bodily meaning that informs gesture, the biomechanical analysis by Pouw and Fuchs (2022) relates articulatory physiology with prosody by asking 'how can gestures reach the vocal system?' (p. 3). Their answer helps to see that a gesture with rotating humerus reverberates beyond the upper limb segments spelled out by Boutet, as it involves chest and back muscles that 'will constrain the rib cage' and 'can affect respiratory functions' (Pouw & Fuchs, 2022, p. 3). By indicating that the physical impulse creating these constraints and affects for the lateral sweep occurs with the speaker's accented 'no', the demonstration also implies acoustic consequences of the lateral sweep for vocalization of 'no'. This link was not made explicit, however, nor were consequences of the lateral sweep for other systems, such as the superficial musculoaponeurotic system that pulls facial expressions.

## 2.6 Current study

To address the gaps and issues raised by the foregoing overview, the current study identified multimodal negation as an ecological context in which vocal-entangled gesture can be inferred. The target data are utterances that involve the recognizable lateral sweep gesture and an accented negative item. The analysis to be performed involves examining selected utterances in visualization software (ELAN and PRAAT) to identify how the kinesthetic and acoustic dynamics correspond as the accented word and gesture materialize. The following section describes how such utterances were discovered and selected in order to make a study of vocal-entangled gesture with qualitative methods and ecological data feasible.

## 3 Methods

### 3.1 Database of negative utterances

The phenomenon to be described was discovered in my database of negative utterances. This is being systematically constructed by trawling popular English-language TV shows, drama series, film excerpts, stand-up comedy, podcasts, debates, academic lectures and other viral video clips that circulate on social media platforms like YouTube, Facebook and WeChat (including pranks, social experiments and street interviews). When examples of utterances with the characteristic linguistic and gestural patterns associated with negation described in Section 2 catch my attention, I flag them up for further analysis by creating a new entry in my Excel log of potentially valuable examples. The motivation for building a database from the genre of online videos is the immediate access to a frequency and diversity of examples that have the potential to accelerate and expand our understanding of gestures associated with negation.<sup>7</sup> A case in point is the sub-corpus of 50 instances where the speaker lengthens the negative particle's onset consonant while gesturing. Preliminary

<sup>7</sup>My experience resonates with BBC journalist Gill's description of online video sharing platforms as offering scientists a 'seemingly endless video library' full of potentially 'scientifically-significant moments' that might otherwise require 'years of intensive fieldwork' to come across (Gill, 2022).

analyses of these instances provided the methodological development and empirical sampling pool for this article's in-depth study.

### 3.2 Selection of examples

In order to analyse onset lengthening of pitch-accented negative items as vocal-entangled gestures systematically, eight examples from the wider corpus were selected. Their lexico-grammatical, acoustic, biomechanical and kinesthetic profiles are similar, but they were uttered by different speakers in different contexts. More specifically, the utterances have the negation words 'never' or 'no' ('standard negation' or 'nuclear negation'). Their particle is accented with saliently lengthened onset (average 300 ms).<sup>8</sup> The speakers all perform a variant of lateral sweeping gesture (pronation/adduction of the hand with external–internal rotation of the humerus) while undergoing some degree of facial deformation. According to descriptors proposed by Rohrer et al. (2023), all selected gestures would be attributed the 'pragmatic function' of 'operational marking': they are 'gestures that operate in conjunction with what is being expressed verbally' (p. 47), with assessment of pragmatic strength ('the gesture's "emphatic" nature or "force"') being 'strong' (p. 51).<sup>9</sup>

The choice to present several contextualized examples that illustrate a recurrent gestural phenomenon identified in a much larger sample is consistent with several related traditions in gesture studies, including context-of-use studies (Kendon, 2004) and recurrent gesture studies (Ladewig, 2024). The general pattern of negative word-onset lengthening with open hand-prone gesture preparation is representative of the other examples in my wider corpus (>n50). However, the gesture–speech biomechanics thesis would predict that different kinesic variants (e.g. with strokes involving oscillatory movements) would affect the vocal-entangled sounds differently. This prediction could be tested in future research.

In a further attempt to strengthen the validity of claims across the selected examples, all were sourced from sub-genres of talk show, reality series and podcast. Notwithstanding the inherent differences across these media events, the examples are homogenous in involving highly expressive and articulate media professionals either addressing an interlocutor in front of cameras (with or without a live studio audience) or addressing the camera directly (monologue). However, exploring whether aspects of meaning and behaviour in my examples are shaped by the specific mediality of these staged speaking situations (e.g. physical layout and camerawork; Luginbühl & Schneider, 2020) was beyond the present study's scope of empirical analysis.

### 3.3 Methods of analysis

After making video clips for each example containing sufficient context, the negative utterances were analysed in ELAN for speech and gesture and PRAAT for acoustics and then synthesized for presentation in the manuscript.

<sup>8</sup>A widely cited average of consonant duration in un-accented English speech is 'About 70 msec' (Klatt, 1976, p. 1209).

<sup>9</sup>Rohrer et al. (2023) use the example of such a 'hand sweep' with 'there's no more brie' to illustrate a case where the pragmatic function of 'operational marking' is 'fairly straightforward' (p. 47).

### 3.3.1 Analysing speech and gesture in ELAN

To study the material in ELAN, a template was created with annotation tiers for speech and gesture. In the tier for speech, broad transcriptions of sounds following the International Phonetic Alphabet were made for each word by listening, slowing down the video and using the waveform to guide me in identifying sound and word boundaries (the speaker's mouthing can be a useful guide too). In the tier for gesture, the linear or 'phrase' structure of the gestures was annotated following Kendon (2004), chapter 7). These include preparation, stroke and retraction phases as well as any holds. Transitions between these phases can be identified by taking the still image in a frame-by-frame analysis (Bressem et al., 2013). For all stroke phases, the 'manual apex' could also be identified as 'points of maximum extension, sudden stops, or changes in direction' (Rohrer et al., 2023, p. 28). Since the 'lateral sweep' gestures in the current study all involve a 'unidirectional stroke' (Rohrer et al., 2023, p. 28), the right-most boundary of the stroke annotation in ELAN was assumed to be the apex. Based on diagrams in research by Pouw and Fuchs (2022), I assumed this manual apex was the lateral sweep gesture's moment of 'physical impulse' or 'peak impetus' (PI). In kinematic terms, this PI is when the momentum of gestural impulse is at its highest (Pouw, de Jonge-Hoekstra, et al., 2020a), which corresponds with 'mechanical loadings of the upper limb onto the body' (Pouw, Harrison, et al., 2020b, p. 1235). Since it is 'during the maximum extension moment when the hand suddenly stops, i.e., when there is a "pulse" in the movement' that 'unintended vocal inflections occur', the acoustic feed corresponding with the endpoint of the lateral sweep could be manually checked for any such inflections (right boundary of stroke annotation in ELAN). Furthermore, since any 'change in momentum of... body segment' potentially 'yields a quantity of force' with implications for bodily systems, this means that moments of acceleration also induce mechanical loadings (Pouw & Fuchs, 2022, p. 17), so left boundary of stroke annotation/right boundary of preparation phase was examined too. Observations of other bodily motions were also made, especially facial distortion.

### 3.3.2 Acoustic analyses in PRAAT

A detailed analysis of acoustic properties such as duration, pitch and intensity readings was carried out in PRAAT. The spectrogram for each example shows visible changes in spectral energy associated with each sound and its wave form over the course of the utterance. Measurements for duration, F0 and intensity contours were taken from within the onset consonant from first syllable of negation words. Guided by details of the spectrogram, selection was made in PRAAT window manually (therefore approximately) from onset of consonant (/n/) to onset of vowel (/ɛ/ or /oʊ/). Automatic pitch and intensity queries were then run to retrieve their maximum and minimum values over the selected period (i.e. the lengthened onset consonant /n/).

### 3.3.3 Correspondence and presentation

Correspondence of the acoustic characteristics with the gesture phrase structure could then be quantitatively and qualitatively examined. I focused on how the acoustics of the initial lengthening of the negative item's consonant corresponds with other bodily motions, especially forelimb gesture preparation and facial

distortion. To present these analyses in the manuscript, tables are first used to present the quantitative data. A PowerPoint slide was then used for their qualitative contextualization as follows. For each example, I copied the utterance's F0 and intensity contours from PRAAT, aligned this with the transcription of sounds in the International Phonetic Alphabet, and then highlighted the onset–consonant lengthening. I added the gesture phase coding from ELAN to align symbols representing the different phases of gestural action again following Kendon (2004): preparation (~ ~ ~), stroke (\*\*\*) , retraction (.-.-) and holds of any of these phases can be underlined (e.g. pre-stroke hold ~ ~ ~). The different bodily transformations under scrutiny around the PI are illustrated with an image from the transition into and out of the stroke (grabbed from left and right boundary of the stroke annotation in ELAN). To show the wider discursive context of each negative utterance, a portion of transcript derived from the machine-generated subtitles on YouTube then reworked for accuracy is presented. However, detailed semantic–pragmatic analyses of the examples were beyond the scope of the current paper. As ELAN does not provide kinematics, the relation between PRAAT's acoustics and physical details of gestural impulse (e.g. rate of acceleration) for the present corpus of examples also awaits future research.

#### 4. Examples and analyses

This section first presents the quantitative results for eight examples (Section 4.1). These measures and values are then contextualized for six examples by presenting qualitative analyses of their entanglement, understood as the correspondence of vocal–respiratory, forelimb motion and superficial musculoaponeurotic systems (Section 4.2).

##### 4.1 Quantitative results

Table 1 presents acoustic measurements of duration (ms), pitch (Hz) and intensity (dB) during initial lengthening of the voiced alveolar nasal consonant /n/ for each of

**Table 1.** Acoustic measurements of duration (ms), pitch (Hz) and intensity (dB) during initial lengthening of the voiced alveolar nasal consonant /n/

Utterance	Duration of /n/ (ms)	Pitch rise (Hz)	Intensity rise (dB)
1 It's <u>never</u> been said	235	77.5 (163.2–240.7)	7.4 (65.3–72.7)
2 <u>Trump never</u> had any evidence of fraud	194	27.1 (132.7–159.8)	5.7 (61.7–67.4)
3 This you may <u>never</u> have heard before	317	37.2 (80.6–117.8)	5.7 (64.6–70.3)
4 Almost <u>never</u> has a party voted	351	16.8 (195.3–212.1)	15.9 (68.8–84.7)
5 <u>Nobody</u> laughs	434	51.2 (110.3–161.5)	21 (52.5–73.5)
6 <u>Literally no one</u> knows what happens now	340	244.6 (84.7–329.3)	26.3 (47–73.3)
7 No one with any level of a brain	297	122.3 (193.3–315.6)	7.9 (56.6–64.5)
8 There's <u>no</u> risk to try this out for yourself	238	82 (98.5–180.5)	13.9 (60.5–74.4)
Total average	300.75	82.3	13



**Table 2.** Average duration, pitch rise and intensity rise during onset consonant of negation word

Vowel length	Duration of /n/ (ms)	Pitch rise (Hz)	Intensity rise (dB)
Short vowel /ε/	274	39.7	8.7
Diphthong /ou/	327.25	125	17.3

the eight examples. The statistics demonstrate that the lengthened consonants also have a pitch and intensity rise.

Contrary to predictions based on Niebuhr's (2010) phonetic profiles of negative emphasis, the duration of lengthened onset consonants for the current set of examples was shorter when followed by a short vowel (/ε/) and longer when followed by the longer diphthong (/ou/). As shown in Table 2, the average onset consonant lengthening of the first syllable of 'never' (/n/ preceding short vowel /ε/) was 274 ms, while the average onset consonant lengthening of 'no' (/n/ preceding diphthong /ou/) was 327.3 ms. Average pitch and intensity rises were also greater when lengthening was followed by the diphthong.

The next section will contextualize these measures and values by exploring their correspondence with other aspects of the speakers' bodily motion, especially forelimb gesturing and facial distortion.

#### 4.2 Qualitative analyses of vocal-entangled gesture at accented negative item onset

The qualitative analysis of examples serves to show that in all the selected utterances to some extent as syllable-onset consonant is lengthening with vocal fold vibration and pulmonic flow increasing (NNNNNN), the speaker's humerus is rotating with hand pronating, adducting and extending, while face is deforming. Two sets of examples will be presented, the first involving utterances with the negative particle 'never' and the second with the particle 'no'. This organization explores body motion's entanglement in two different acoustic environments of the accented syllable, offering three examples from each environment.

##### 4.2.1 Examples with clausal negation 'never' (/n/ precedes short mid-front unrounded vowel /ε/)

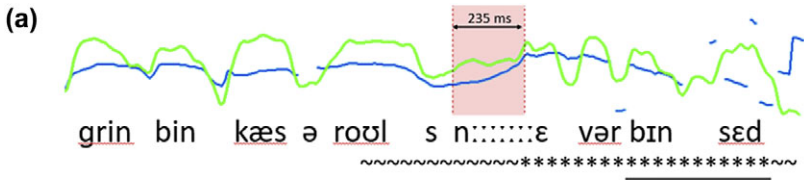
For example, lexico-grammatical pattern is clausal negation, specifically 'standard negation' ('negation of declarative main clauses with a verbal predicate'; Miestamo, 2017, p. 406). As forelimb gesture is preparing in examples with 'never', consonant is lengthening (with pitch and intensity both rising), cheeks are raising, lips are stretching and eyelids are closing/squinting.

##### **Example 1** (It's) never been said.<sup>10</sup>

Our first example was discovered during the opening monologue to an episode of *Jimmy Kimmel Live!* in the context of a joke about Thanksgiving side dishes. The utterance '(it) has never been said' is the host Jimmy's answer to his rhetorical

<sup>10</sup>Posted to the official *Jimmy Kimmel Live!* YouTube channel on 22 November 2023 (2.5 M views to date). <https://www.youtube.com/watch?v=dvUYZDNjd1o&t=160s> (02.30).

question ‘ever heard anyone say pass the green bean casserole?’ (Transcript 1, line 2). In terms of the overall prosodic profile visible in the PRAAT window, the expression ‘never been’ exhibits a plateau-like shape F0 contour with clear intensity peaks on each syllable. Slope of F0 (77.5 Hz) and first intensity peak (7.4 dB) start rising during lengthening of syllable–onset consonant /n/ (235 ms). This corresponds with the speaker’s internal rotation of the humerus (manual gesture preparation phase) as cheeks are raising, lips are stretching and eyelids are closing/squinting (Figure 2a). It turns out to be preparation for a unidirectional



**Figure 2.** (a) Image from left boundary of stroke phase (internal medial rotation). (b) Image from right boundary of stroke phase (external lateral rotation).

**Table 3.** Acoustic measurements for Example 1

Ex	Utterance	Duration of /n/ (ms)	Pitch rise (Hz)	Intensity rise (dB)
1	(it's) <u>never</u> been said	235	77.5 (163.2–240.7)	7.4 (65.3–72.7)

gesture stroke beginning with the vowel onset (/ɛ/). As '(n)ever' is uttered, the stroke involves external rotation of the humerus ('lateral sweep'; Figure 2b). This stroke reaches its apex (endpoint) with 'been' (which has an intensity peak) and is held with 'been said'. Jimmy's face undergoes changes during this part of utterance due to mouthing constraints from vocalizing but does not fully relax until completion of post-stroke hold (eyes and lips maintain degrees of squinting and stretching). Zooming into the F0 and intensity contours, bobbles become visible at onset (max acceleration) and offset (max deceleration) of gesture stroke (Table 3).

### Transcript 1

1. even the people who like green bean casserole do not like green bean casserole
2. ever heard anybody say pass the green bean casserole (it's) never been said
3. I had an argument though in the office today

**Example 2.** Trump never had any evidence of fraud.<sup>11</sup>

The second example 'Trump never had any evidence of fraud' was sourced from the 'Closer Look' monologue segment of *Late Night with Seth Meyers*. This utterance appears in a subordinate *that*-clause as the direct object of the sentence 'Powell essentially admitted to prosecutors' (Transcript 2, line 2). In this acoustic profile, 'never had' exhibits a plateau-like shape F0 peak with clear intensity peaks on each syllable. Syllable-onset consonant /n/ is lengthened for 235 ms, during which time both F0 and intensity rise (respectively, 27.1 Hz and 5.7 dB). As in the first example, it is when lengthening syllable-onset consonant with increased vocal fold vibration and pulmonic flow that speaker is rotating the humerus (internal) of both hands while his cheeks are raising, lips are stretching and eyelids are closing/squinting (Figure 3a). Uttering of '(n)ever had' likewise involves humerus rotating (now external) without face relaxing (though changing due to mouthing) until completion of stroke at onset of 'had' and instant retraction (Figure 3b). No acoustic bobble can be seen at the endpoint of stroke, which smoothly transitions to rest without any post-stroke hold (Table 4).

### Transcript 2

1. for example Powell essentially admitted to prosecutors
2. that Trump never had any evidence of fraud
3. nor did he even attempt to provide any evidence of fraud
4. except for when he claimed he saw something weird on his TV

<sup>11</sup>Posted to the official *Late Night with Seth Meyers* YouTube channel on 17 November 2023 (2.4 M views to date). <https://www.youtube.com/watch?v=PdMUcQajXxg> (08.31).

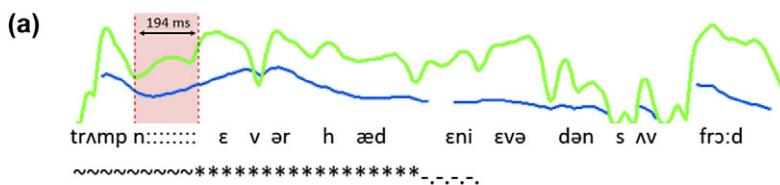


Figure 3. (a) Image from left boundary of stroke phase. (b) Image from right boundary of stroke phase.

Table 4. Acoustic measurements for Example 2

Utterance	Duration of /n/ (ms)	Pitch rise (Hz)	Intensity rise (dB)
2 Trump <u>never</u> had any evidence of fraud	194	27.1 (132.7–159.8)	5.7 (61.7–67.4)

**Example 3:** This you may never have heard before.<sup>12</sup>

Our third example is this article's title track, the utterance 'this you may never have heard before' from an extract of the Television Show *Celebs Go Dating*.<sup>13</sup> The dating agent Paul has planned to 'call out' his client Miles (from *Made in Chelsea*) for 'acting with his dates'. Seated opposite Miles, he is going to say 'I believe you are selfish' (Transcript 1, line 5). To soften this face-threatening call out, Paul first says 'this you may never have heard before' with a plateau-like shape F0 peak over 'never have heard' involving intensity peaks on each syllable (line 4). With focus-fronted 'this', Paul has a 'precision grip' gesture (Kendon, 2004, Ch. 12). This is from where a palm down gesture emerges whose stroke occurs with 'never'. While saying 'you may', thus, Paul's hands are pronating, abducting and extending with some internal rotating of the humerus. This continues while lengthening the negative syllable-onset consonant /n/ (317 ms) with F0 and intensity increasing (respectively, 37.2 Hz / 5.7 dB), while cheeks are raising, lips are stretching and eyelids are closing/squinting (Figure 4a). Acceleration of hands into the gesture stroke begin with vowel onset, as Paul's humerus rotates externally and facial musculature relaxes (Figure 4b). Bobbles in pitch contour are visible at onset (max acceleration) and offset (max deceleration) of gesture stroke (Table 5).

### Transcript 3.

1. Paul: there's something *i* have wanted to tell you miles
2. Miles: oh god yeah
3. Paul: here's the deal here's the deal
4. this you may never have heard before
5. I am going to tell you something
6. this is from the bottom of my heart
7. i believe you are selfish

#### 4.2.2 Examples with 'no' (/n/ precedes diphthong /ou/ starting with mid-back rounded vowel /o/)

The next set of examples involves the negative item 'no' with its diphthong /ou/. On average, these examples show comparatively greater lengthening of onset consonant (avg. 337.3) with higher rises in F0 and intensity during the lengthening (avg. 125 Hz and 173 dB, respectively). They allow us to see different facial distortion during onset lengthening with forelimb gesture preparing, one source of which appears to be articulatory rounding of the negative particle's vowel. In these examples, more specifically, as syllable-onset consonant is lengthening with vocal fold vibration and pulmonic flow increasing, the speaker's humerus is rotating with hand pronating and adducting. Lips are funnelling, while eyelids are raising (as sometimes are

<sup>12</sup>Posted to the E4 YouTube channel in early 2022 (56 K views to date) <https://www.youtube.com/watch?v=kHL8qRH1vmA> (01:25).

<sup>13</sup>This utterance may be classified linguistically as having clausal negation because the adverb 'never' negates the whole clause (of the theoretical sentence 'you may have heard [this] before'). The sentence therefore also demonstrates front focus.

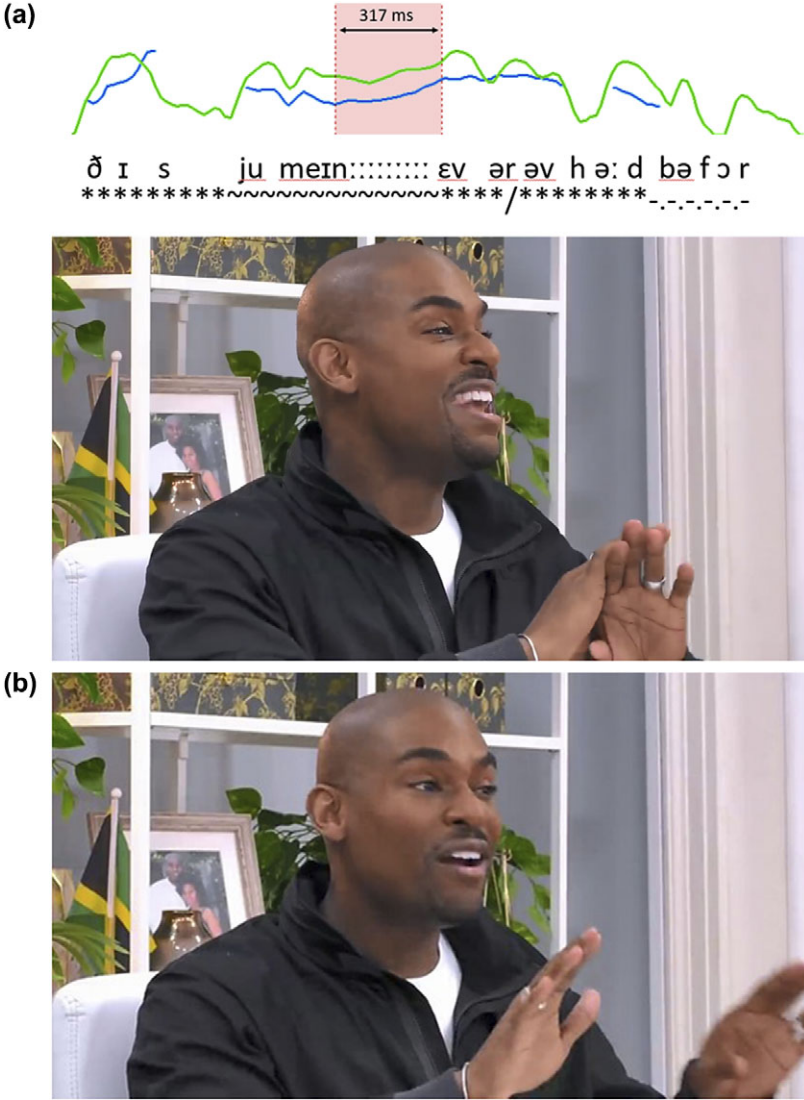


Figure 4. (a) Image from left boundary of stroke phase (i.e. right boundary of pre-stroke hold). (b) Image from right boundary of stroke phase (i.e. left boundary of second stroke phase).

Table 5. Acoustic measurements for Example 3

	Utterance	Duration of /n/ (ms)	Pitch rise (Hz)	Intensity rise (dB)
3	This you may <u>never</u> have heard before	317	37.2 (80.6–117.8)	5.7 (64.6–70.3)

**Table 6.** Acoustic measurements for Example 4

Ex	Utterance	Duration (ms)	Pitch rise (Hz)	Intensity (dB)
5a	<u>Nobody</u> laughs	434	51.2 (110.3–161.5)	21 (52.5–73.5)
5b	<u>Nobody</u>	222	63.6 (107.5–171.1)	18.7(53.7–72.4)

eyebrows). In terms of lexico-grammatical pattern, this second set of examples is negative by virtue of a nuclear negative word in subject position (Examples 5–7).

### Example 5. Nobody laughs.

Our first example of this scenario was perceived while enjoying a widely viewed video on YouTube called Jamie Foxx Roasted Mike Tyson to His Face from the Foxx's interview on *The Tonight Show Starring Jimmy Fallon*.<sup>14</sup> The segment of interest begins with Foxx setting the scene by saying 'so I'm in the hood and I'm killing LA and I get to my Mike Tyson joke and that's where I usually like get a standing ovation' (lines 1–4). He now says 'and when I get to the joke' (line 5), pauses and then says 'nobody laughs' (line 6). The audience laugh (line 7), then Foxx repeats 'nobody' (line 8) and asks rhetorically 'you know why?' (line 9). He answers 'because Mike Tyson is in the building' (line 10). The acoustic measurements for Foxx's two 'nobody' are shown in Table 6.

Focusing on the first instance, Foxx's gesture stroke is timed with the diphthong /ou/ of 'nobody' (instance 5a, Table 6). This first syllable's rime also sees pitch and intensity peaks. Its onset involves a lengthened buzz of the voiced alveolar nasal consonant /n/. At nearly half a second duration (434 ms), this is the longest one in the current set of examples. This buzz enfolds an F0 rise of 51.2 Hz and intensity rise of 21 dB, which are also relatively large (cf. Table 1). This intensified and pitched sound lengthening is prefaced with an audible inhalation (.hhhh). These vocal-tract behaviours correspond with the preparation phase of a two-handed 'lateral sweep' gesture and facial activity involving lip rounding (Figure 5a). Foxx's whole upper body is therefore inflating and lifting, suggesting evidence for the 'tight relations between respiration-related activity and vocalization' (Pouw, de Jonge-Hoekstra, et al., 2020a, p. 1231). Chest circumference is included among the 'respiratory kinematic changes' known to be 'amplified by upper body limb movement' (Pouw & Fuchs, 2022, p. 5). Gestural stroke is during the syllable's rime (the diphthong /ou/) while facial distortion maintains rounding or 'funneling' of the lips and eyelids raising (Figure 5b). Note these are different facial distortions from when the syllable rime was the /ε/ of 'never'.

**Transcript 5.** 1 so I'm in the hood and I'm doing this joke  
 2 and I'm killing LA (2ZP)  
 3 and I get to my Mike Tyson joke and  
 4 that's where I usually like get a standing ovation  
 5 and when I get to the joke  
 6 .hhhhnobody laughs

<sup>14</sup>Posted to *The Tonight Show Starring Jimmy Fallon* YouTube channel on 26 May 2017 (7.9 M views to date). <https://www.youtube.com/watch?v=ssRaMYMmis8> (01:25).



Figure 5. (a) Image from right boundary of gesture preparation phase qua left boundary of stroke phase. (b) Image from right boundary of gesture stroke phase qua left boundary of post-stroke hold phase.

- 7 (audience laughter)
- 8 nobody
- 9 you know why
- 10 because Mike Tyson is in the building

Comparing Foxx’s two instances of ‘nobody’, they both have initial lengthening. There is no in-breath with the repetition, which is only half as long (222 ms compared to 434 ms), yet this repeated, shorter ‘nobody’ has a steeper pitch rise (63.6 Hz). They have similar changes in intensity or volume (difference of 2.3 dB). The total duration of the first ‘nobody’ is 860 ms, which means the lengthening of onset accounts for



50%. The second nobody is 691 ms, which means the lengthening of onset accounts for 32%.

**Example 6.** Literally no one.<sup>15</sup>

Our next example with ‘no’ is from an episode of *The Late Show with Stephen Colbert* called Kevin McCarthy’s Groundhog Day. Colbert opens his monologue by satirising American politician Kevin McCarthy’s failed attempts to be elected as speaker of the house. After turning to his set hand to check ‘the house did they fail again?’ (line 1), Colbert addresses the camera to say ‘no one, no one, literally no one’ (lines 2–4), parenthetically remarking ‘and I called people who know things’ (line 5), then completing his utterance ‘no one knows what happens now’ (line 6). Each ‘no one’ is timed with a repeated stroke of one-handed ‘lateral sweep’ gestures (Figure 6a–c). Table 7 includes specific measures of duration, pitch and intensity for the negative item receiving greatest accentuation, while Figure 6a montages data from ELAN and PRAAT to visualize acoustic and kinesthetic phenomena. Duration of the voiced alveolar nasal consonant /n/ increases with each repetition: 115, 141 and 340 ms. Increases in F0 and intensity are similarly commensurate, the third instance having both the biggest pitch rise (244.6 Hz) and biggest intensity rise (26.3 dB) in this study’s selection of utterances. In this example, thus, the longer the duration, the bigger the pitch rise. Gesturally, the third instance also has a marked preparation phase corresponding with the modifier ‘literally’. As can be seen from the screenshots (Figure 6a–c), Colbert’s lips are funnelling while eyelids and brows are raising.

**Transcript 6.**

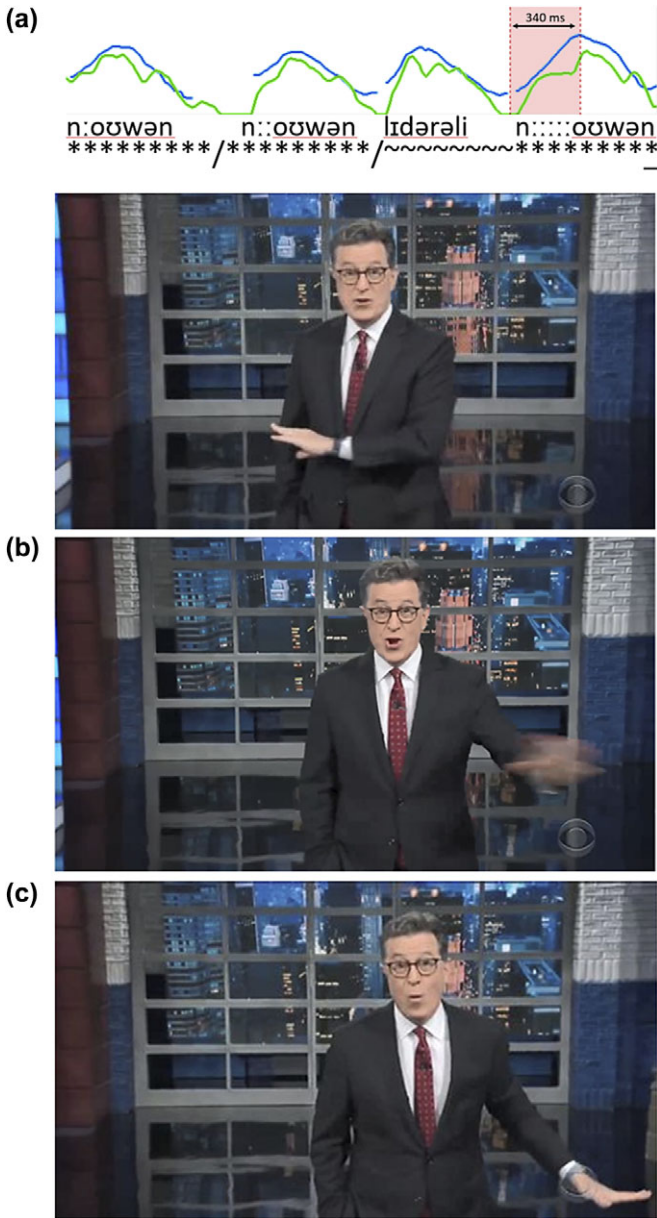
- 1 the house did they fail again okay good
- 2 no one
- 3 no one
- 4 literally no one
- 5 and I called people who know things
- 6 no one knows what happens now
- 7 here’s what we do know
- 8 it’s fantastic

**Example 7.** No one with any level of a brain.<sup>16</sup>

In this video clip found circulating on YouTube Shorts, online personality Emily Wilson shares her perception of women who earn money by trading explicit videos of themselves on sites such as Only Fans. Speaking about the longer-term relationship prospects for young adult content creators, EW’s concern is that ‘no one with any level of like a brain would ever respect somebody who does that’ (lines 3–4). As her syllable–onset consonant /n/ is lengthening (297 ms), F0 and intensity are increasing (122.3 Hz/7.9 dB) (Table 8). Both the speaker’s humerus bones are rotating internally with hands pronating and adducting, as her face is deforming with lips funnelling and brows lowering (Figure 7, image on left). This is preparation for a lateral sweep

<sup>15</sup>Posted to The Late Show with Stephen Colbert YouTube channel on 5 January 2023 (2.1 M views to date). [https://www.youtube.com/watch?v=XZ\\_DuHqyZs8&t=30s](https://www.youtube.com/watch?v=XZ_DuHqyZs8&t=30s) (0:14).

<sup>16</sup>Originally sampled from Jamie Kennedy’s Podcast with Emily Wilson as guest uploaded to YouTube on 2 June 2023 (70.3 K subscribers; 5,168 views to date) <https://www.youtube.com/shorts/o4fnlgEXv6E>.



**Figure 6.** (a) Image from right boundary of gesture preparation phase qua left boundary of stroke phase. (b) Image from stroke phase. (c) Image from right boundary of stroke phase qua left boundary of post-stroke hold phase.

variant involving both hands (referred to elsewhere as 2PDmid; Harrison, 2018, pp. 33–39), which in this instance also exhibits some ‘recoil’ at stroke apex (Figure 7, image on right; Rohrer et al., 2023, p. 28). EW holds her open hand-prone gesture

Table 7. Acoustic measurements for Example 6

Utterance	Duration of /n/ (ms)	Pitch rise (Hz)	Intensity rise (dB)
6 Literally <u>no one</u> (knows what happens now)	340	244.6 (84.7–329.3)	26.3 (47–73.3)

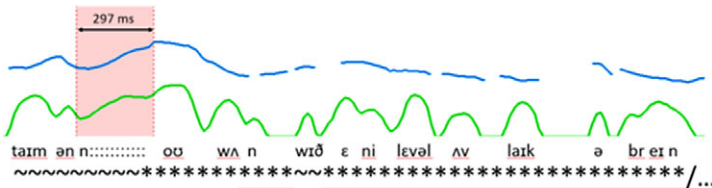


Figure 7. Image from left and right boundary of stroke phase (before hold).

momentarily before preparing and holding palm-up gestures with the rest of her utterance.

**Transcript 7.**

- 1 You're going to hit the real world eventually
- 2 It's just a matter of time
- 3 And no one with any level of like a brain
- 4 Would ever respect somebody who does that

**Table 8.** Acoustic measurements for Example 7

	Utterance	Duration of /n/ (ms)	Pitch rise (Hz)	Intensity rise (dB)
7	No one with any level of a brain	297	122.3 (193.3–315.6)	7.9 (56.6–64.5)

## 5. Discussion

Movement scientists have proposed to ground the relation between prosody and gesture in ‘vocal-entangled gestures’ (Pouw & Fuchs, 2022), defined as ‘bodily utterances which can have consequences for respiration and vocalisation’ (p. 13), opening up new ways to look at phenomena more commonly described as multimodal. By analysing the acoustic signal and gestural movements in the domain of multimodal negation (accented negation words, co-occurring lateral sweep gestures and facial expressions), this article identified an acoustic profile with which gestural form and organization are plausibly entangled. Biomechanical linkages between upper limb movement and the respiratory–vocal system (e.g. involving connective muscle tissues) provide a new perspective on the articulatory behaviour of onset–consonant lengthening with forelimb gesture preparation discovered in the present research. The perspective opened by ‘vocal-entangled gestures’ enriches our current understanding of negative utterances grounded in psycholinguistic, cognitive–linguistic, communicative–semiotic and developmental research, linking previous work on multimodal negation to enactive–ecological models of whole-bodily languaging.

The case was made with eight examples selected from fifty instances of lengthening and examined in ELAN/PRAAT. The analyses show that as syllable–onset consonant is lengthening (voiced alveolar /n/ = 300 ms on average) with vocal fold vibration and pulmonic flow increasing (respectively, by 82.3 Hz and 13 dB on average), the speaker’s humerus is rotating with hand pronating and adducting (forelimb gesture preparation of a ‘lateral sweep’), while his or her face is deforming. These key findings can be discussed with respect to different strands of empirical research and different theories of utterance.

### 5.1 Word-onset lengthening as emphasis of intensity and bodily meaning–feeling

Lengthening of the voiced alveolar nasal /n/ at the onset of negative items is consistent with the form of semantic–pragmatic emphasis called ‘negative intensification’ (Niebuhr, 2010). This kind of lengthening intensifies ‘a negative valence’ (Niebuhr, 2010, p. 173), which associates the lengthening of the consonant with ‘expressive and attitudinal aspects of [negative] meaning’ (p. 195).<sup>17</sup> In referring to Niebuhr’s work, Ward (2019) subsequently includes lengthening under ‘iconic uses’ (p. 57) of prosody. Ward lists ‘negative sentiment’ as its ‘representative meaning or function’ in his typology of ‘likely primordial prosody–meaning mappings’ (Ward, 2019, pp. 57–60). The current study makes the intuitive link between negative

<sup>17</sup>Valence refers to the positive or negative attractiveness of the emotion’, with examples of ‘negatively valenced emotions’ being ‘anger, fear, and anxiety’ (Thibault, 2021, p. 189).

sentiment inherent to sound lengthening and the lexico-grammatical manifestation of negation. The analyses indicate that the onset lengthening of accented negative words in English (voiced alveolar nasal consonant /n/) is plausibly entangled with other bodily gestures from physiologically related systems, including ‘lateral sweep’ forelimb gestures and facial distortion.

In making this link, the current study also brings new insights into theories of language that recognize environmentally embedded bodily movement as the common denominator of all linguistic interactive behaviour and accordingly begin with the entanglement of speech and gesture (Di Paolo et al., 2018; Harrison, in preparation; Pouw & Fuchs, 2022; Thibault, 2021). In Thibault’s relational ontology of languaging, for instance, the importance of word-onset behaviour plays a key role in articulating the interplay between different scales of language, namely, between first-order embodied interactivity and second-order lexico-grammar (Thibault, 2021, Ch. 4). Indeed, the /n/ of negation words can be recognized as a ‘gesture-sound complex’ or ‘submorphemic marker’. It illustrates the go-between of ‘time-extended ecological activities’ on the pico-scale (like the voicing, tongue movement and nasalization of /n/) and ‘meta-linguistic objects’ on the cultural–historical scale (like a grammar–gesture nexus of negation).<sup>18</sup> From a whole-bodily languaging perspective, this paper’s examples of consonant lengthening show speakers actively exploring and manipulating the sensory–kinetic experience (bodily feeling and its associations – especially negative valence) said to be ‘embodied’ by such articulatory events.<sup>19</sup> They also show how the importance conferred to the onset of words in languaging theory naturally extends to the onset of other bodily gestures, namely gesture preparation and facial distortion. The implicit deixis of submorphemic markers mentioned by Thibault (2021) – their ‘capacity to evoke kinaesthetic memory of sensory-kinetic experience’ (p. 207) – is consistent with the theory of vocal-entangled gestures. Such gestures are argued to ‘have communicative potential by serving as an index for some embodied state of affairs rather than as a representation of purely mental content’ (Pouw & Fuchs, 2022, p. 1). Negation’s bodily state of affairs, I propose, includes a negatively valenced feeling induced (and shared) from extended vocal fold vibration of voiced alveolar /n/. From an entangled perspective on utterances, such resonant feeling should be included among the other sources proposed for the lateral sweep’s negative meaning (cf., Boutet, 2015, 2018; Bressemer & Müller, 2014; Calbris, 2011). It could likewise be entertained as a way people understand each other’s negation, on a scale that needs reconciling with the unfolding context of interaction (Beaupeil-Hourdel & Morgenstern, 2021) and participants’ pragmatic inferencing (Prieto et al., 2013).

The precise ratio of lengthening of consonant duration vis-à-vis vowel length in my examples was found to be reversed to Niebuhr’s (2010) examples (i.e. positive not negative). Further research would be needed to validate this difference and identify its source among numerous possibilities, including the different languages (English/German), target syllables (Niebuhr did not include negation words), empirical

<sup>18</sup>For the submorphemic link between /n/ and negation, see Bottineau’s theory of cognemes on which Thibault draws (Bottineau, 2007, p. 57).

<sup>19</sup>‘Embodied’ deserves scare quotes here because it risks implying that bodily motions (whether of vocal-tract, forelimb or face) somehow encase negative valence, whereas from enactive–ecological languaging perspectives these synergies of bodily motions *are* the negative valence ‘working through’ the speaker (cf. Thibault, 2021, p.186).

methods, sample size and speech genre. Similarly requiring future research are the acoustics of negation words entangled with different kinesic variants of open hand-prone gestures (e.g. when involving oscillatory movements). While the lengthening of onset consonant during gesture preparation is representative of the wider corpus (>n50), the gesture–speech biomechanics thesis would predict that variations in kinesic form will affect the respiratory system differently. Predictions could also be formulated with a view to comparing negation across typologically different languages.

### 5.2 Not the ‘Not Face’?

The findings of this study also revealed the potential implication in vocal-entangled gestures of physiological systems yet to be studied in the cited gesture–speech physics research, especially the superficial musculoaponeurotic system (a source of facial expression). Examples were selected to allow comparison of accented syllables whose onset consonant was the same (voiced alveolar nasal /n/) but whose rimes were different (either short unrounded front vowel /ɛ/ or rounded diphthong /ou/). Analyses showed that during lengthening of onset while forelimb gesture is preparing, speakers’ faces were distorting differently with respect to the phonetic profile of the syllable’s rime, most notably vowel rounding. In examples involving ‘never’ (/n/ precedes short mid-front unrounded vowel /ɛ/), lips are raising and eyelids are closing during onset lengthening. In examples with ‘no’ (/n/ precedes diphthong /ou/ starting with mid-back rounded vowel /o/), lips are funnelling and eyelids (as well as sometimes brows) are raising. These differences lead to distinct facial expressions at the distortion apex (Figures 8 and 9).

This study’s examples correspond to what Benitez-Quiroz and team (Benitez-Quiroz et al., 2016) would call negative facial expressions being ‘used... as a co-articulator in negative sentences in spoken languages’ (p. 78). Yet facial distortions with these utterances are different to those described by Benitez-Quiroz and team. Neither sets of examples resemble what they have influentially designated the ‘not face’ and defined as ‘a facial expression of negation that is produced by using the same AUs (Action Units) by people of different cultures’ (p. 82). The Ekmanian ‘action units’ identified for the ‘not face’ are absent from my examples, and my examples involve action units that were not in the combination described for the ‘not face’, such as squinting/blinking and upper lip raising. In my findings, moreover, features of facial distortion at syllable onset are clearly related to the phonetic details of the vowel they precede (namely rounding), whereas Benitez-Quiroz and team ‘did not find any



**Figure 8.** Facial distortion during lengthened /n/ preceding short mid-front unrounded vowel /ɛ/ (lips are raising and eyelids are squinting/closing).



**Figure 9.** Facial distortion during lengthened /n/ preceding rounded diphthong /ou/ (lips are funnelling and eyelids are raising).

difference in the production of the ‘not face’ as a function of the sentence, clause or response’ (p. 81). Since their descriptions of the ‘not-face’ with spoken language negation were based on a small sample (only twelve instances of the ‘not face’ appearing with negative sentences, eight of which were in English), their claim to have ‘identified a facial expression of negation that is consistently used... in language as a co-articulator’ (p. 82) could seem overstated. Furthermore, this claim was presented as key evidence for a theory of language evolution in which negative facial expressions of emotion were an evolutionary precursor to facial expressions that ‘co-articulate’ with negative sentences in spoken language (Benitez-Quiroz et al., 2016). Though similarly based on a small sample, the close relations between facial distortions and phonetic details found in the current study would be more consistent with an evolutionary picture to which the entanglement of bodily motion with vocalization is central (as articulated, for example, by Pouw & Fuchs, 2022). Future bio-mechanical experiments informed by the relevant anatomical facts and using appropriate sensing technologies to perceive physical entanglements of the different systems would help to strengthen such findings.

## 6 Conclusion

Multimodal and entangled treatments of the relation between speech and gesture yield fundamentally different perspectives on the utterance with important implications for gesture studies. Whether we see an ensemble of semiotic components that the speaker creatively orchestrates or material bodily dynamics with their own biological animacy depends on our theory and methods. Yet the study of ‘co-speech gesture’ and ‘vocal-entangled gesture’ can be brought together in the analysis of key linguistic phenomena, enriching how we perceive and understand relations between systems of bodily motion. With qualitative analyses of negative sentences sourced from popular televised dialogues, this study concludes that utterances widely

characterized as multimodal negation (comprising lateral sweep gesture, prosodic stress, negative particle and facial expression) plausibly materialize from entanglements of vocal–respiratory, forelimb motion and superficial musculoaponeurotic systems. This materialization process illustrates language’s multiscalarity.

**Data availability statement.** All materials analysed in this article are publicly available on video sharing sites. Web links have been provided for each example.

**Acknowledgements.** I thank the editorial team and anonymous reviewers for detailed comments that improved the manuscript, as well as numerous peers who offered valuable feedback on this work during seminar and conference presentations. I am grateful to my colleagues Alice Chan for help with grammatical classification and Christoph Hafner for discussions of multimodality.

## References

- Beaupoil-Hourdel, P. (2015). *Multimodal acquisition and expression of negation. Analysis of a videotaped and longitudinal corpus of a French and an English mother-child dyad* [Ph.D. Dissertation, Sorbonne Nouvelle University].
- Beaupoil-Hourdel, P., & Morgenstern, A. (2021). French and British children’s shrugs: A crosslinguistic developmental case-study of a recurrent gesture. *Gesture*, 20(2), 180–218.
- Beaupoil-Hourdel, P., Morgenstern, A., & Boutet, D. (2016). A child’s multimodal negations from 1 to 4: The interplay between modalities. In C. Larrivee & C. Lee (Eds.), *Negation and polarity: Experimental perspectives* (pp. 95–122). Springer.
- Benitez-Quiroz, F. C., Wilbur, R. B., & A. M. Martinez (2016). The not face: A grammaticalization of facial expressions of emotion. *Cognition*, 150, 77–84.
- Blanchette, M. N., & Nadeau, M. (2018) Prosody and the meanings of English negative indefinites. *Journal of Pragmatics*, 129, 123–139. <https://doi.org/10.1016/j.pragma.2018.03.020>
- Bottineau, D. (2007). The cognemes of the Spanish language: Towards a cognitive modelization of the submorphemic units in the grammatical words of the Spanish language. *The Public Journal of Semiotics*, 1 (2), 50–74.
- Boutet, D. (2015). Conditions formelles d’une analyse de la négation gestuelle. *Vestnik of Moscow State Linguistic University*, 717, 116–129. [http://www.vestnik-mslu.ru/vestnik.asp?vest\\_lang=Eng&vest\\_type=gum](http://www.vestnik-mslu.ru/vestnik.asp?vest_lang=Eng&vest_type=gum)
- Boutet, D. (2018). Pour une approche kinésiologique de la gestualité. Linguistique, Université de Rouen.
- Boutet, D., Blondel, M., Beaupoil-Hourdel, P., & Morgenstern, A. (2021). A multimodal and kinesiological approach to the development of negation in signing and non-signing children. *Languages and Modalities*, 1, 31–47.
- Bressem, J., Ladewig, S., & Müller, C. (2013). Linguistic annotation system for gestures. In C. Müller, E. Fricke, S. Ladewig, D. McNeill, & S. Tefendorf (Eds.), *Body – language – communication: An international handbook on multimodality in human interaction* (Vol. 1, pp. 1098–1124). De Gruyter Mouton.
- Bressem, J., & Müller, C. (2014). The family of AWAY-gestures. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction* (Handbooks of Linguistics and Communication Science, Vol. 38, no. 2, pp. 1592–1604). De Gruyter.
- Bressem, J., & Müller, C. (2017). The “Negative-Assessment-Construction”—A multimodal pattern based on a recurrent gesture? *Linguistics Vanguard*, 3(s1). <https://doi.org/10.1515/lingvan-2016-0053>.
- Brown, A., & Kamiya, M. (2019). Gesture in contexts of scopal ambiguity: Negation and quantification in English. *Applied Psycholinguistics*, 40(5), 1141–1172. <https://doi.org/10.1017/S014271641900016X>
- Brown, L., & Prieto, P. (2021). Gesture and prosody in multimodal communication. In M. Haugh, D. Kádár, & M. Terkourafi (Eds.), *The Cambridge handbook of sociopragmatics* (Cambridge Handbooks in Language and Linguistics, pp. 430–453). Cambridge University Press. <https://doi.org/10.1017/9781108954105.023>
- Calbris, G. (2005, June). La négation. Son symbolisme physique. In 2ème Conférence ISGS, Lyon (pp. 15–18). <http://gesture-lyon2005.ens-lyon.fr/IMG/pdf/CalbrisFinal.pdf>
- Calbris, G. (2011). *Elements of meaning in gesture*. John Benjamins.



- Cienki, A. (2012). Usage events of spoken language and the symbolic units we (may) abstract from them. In K. Kosecki & J. Badio (Eds.), *Cognitive processes in language* (pp. 149–58). Peter Lang.
- Cuffari, E. (2022). Partial gestures, partial languaging?: Searching for participatory sense-making in video-mediated classrooms. Paper presented at A Day of Sense-Making, ISGS Hong Kong / City University of Hong Kong, Friday, 9 December 2022.
- Darwin, C. (1872). *The expression of emotions in man and the animals*. John Murray
- de Jong, K. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, 97, 491–504.
- Di Paolo, E., Cuffari, E., & De Jaeger, H. (2018). *Linguistic bodies: The continuity between life and language*. MIT Press.
- Esteve-Gibert, N., Prieto, P., & Shattuck-Hufnagel. (this issue).
- Gill, V. (2022). Why scientists are also watching animal YouTube videos. BBC, 29 May 2022.
- Harrison, S. (2010). Evidence for node and scope of negation in coverbal gesture. *Gesture*, 10(1), 29–51.
- Harrison, S. (2014a). The organisation of kinesic ensembles related to negation. *Gesture*, 14(2), 117–141. <https://doi.org/10.1075/gest.14.2.01har>
- Harrison, S. (2014b). Head shakes: Variation in form, function, and cultural distribution of a head movement related to “no”. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & J. Bressemer (Eds.), *Body-language-communication: An international handbook on multimodality in human interaction* (Handbooks of Linguistics and Communication Science, Vol. 38, no. 2, pp. 1496–1501). De Gruyter.
- Harrison, S. (2018). *The impulse to gesture: Where language, minds, and bodies intersect*. Cambridge University Press.
- Harrison, S., & Ladewig, S. (2021). Recurrent gestures throughout bodies, languages, and cultural practices. *Gesture*, 20(2), 153–179.
- Harrison, S. (2024). On Grammar-Gesture Relations: Gestures Associated with Negation. In A. Cienki (Ed.), *The Cambridge Handbook of Gesture Studies* (pp. 446–474). (Cambridge Handbooks in Language and Linguistics). Cambridge University Press. <https://doi.org/10.1017/9781108638869.018>
- Hepp, A., & Couldry, N. (2023). Necessary Entanglements: Reflections on the Role of a “Materialist Phenomenology” in Researching Deep Mediatization and Datafication. *Sociologica*, 17(1), 137–153. <https://doi.org/10.6092/issn.1971-8853/15793>
- Horn, L. (1989). *A natural history of negation*. University of Chicago Press
- Horn, L. R. & Wansing, H. (2022). Negation. In Edward N. Zalta & Uri Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition).
- Huddleston, R. D., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge University Press.
- Huddleston, R., & Pullum, G. K. (2005). *A student's introduction to English grammar*. Cambridge University Press.
- Ingold, T. (2021). *Correspondences*. Polity.
- Jespersen, O. (1917). *Negation in English and other languages*. A. F. Høst.
- Kaufmann, A. (2002). Negation and prosody in British English: a study based on the London–Lund Corpus. *Journal of Pragmatics*, 34(10–11), 1473–1494.
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.
- Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, 59, 1208–1221.
- Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, 57, 396–414.
- Ladewig, S. H. (2024). Recurrent gestures. Cultural, individual, and linguistic dimensions of meaning. In A. Cienki (Ed.), *The Cambridge handbook of gesture studies*. Cambridge University Press.
- Lapaire, J. R. (2006). Negation, reification and manipulation in a cognitive grammar of substance. In S. Bonnefille & S. Salbeyre (Eds.), *La Negation* (pp. 333–349). Presses universitaires François-Rabelais.
- Lapaire, J. R. (2011). Grammar, gesture and cognition: Insights from multimodal utterances and applications for gesture analysis. *Visnyk of the Lviv University. Series Philology*, 52, 88–103.
- Lawler, J. (2005). Negation and NPIs. Available at: <https://websites.umich.edu/~jlawler/NPIs.pdf>.
- Li, F., González-Fuente, S., Prieto, P., & Espinal, M. T. (2016). Is Mandarin Chinese a truth-based language? Rejecting responses to negative assertions and questions. *Frontiers in Psychology*, 7, 1967. <https://doi.org/10.3389/fpsyg.2016.01967>

- Linell, P. (2011/2005). *The written language bias in linguistics: Its origin and transformations*. Routledge.
- Luginbühl, M., & Schneider, J. G. (2020). Medial shaping from the outset: On the mediality of the second presidential debate, 2016. *Journal für Medienlinguistik*, 3(1), 57–93.
- Massumi, B. (2015). *Politics of affect*. John Wiley.
- Miestamo, M. M. P. (2017). Negation. In A. Aikhenvald & R. M. W. Dixon (Eds.), *Cambridge Handbook of linguistic typology* (pp. 405–439). Cambridge University Press.
- Morgenstern, A., Chevrefils, L., Blondel, M., Vincent, C., Thomas, C., Jego, J. F., & Boutet, D. (2021). “Of thee I sing”: An opening to Dominique Boutet’s kinesiological approach to gesture. *Language and Modalities*, 1, 3–16.
- Müller, C. (2017). How recurrent gestures mean: Conventionalized contexts-of-use and embodied motivation. *Gesture*, 16(2), 277–304.
- Niebuhr, O. (2010). On the phonetics of intensifying emphasis in German. *Phonetica*, 67: 170–198.
- Pendleton-Jullian, A., & Brown, J. S. (2023). In search of ontologies of entanglement. *Dædalus*, 152(1), 265–271.
- Pouw, W., de Jonge-Hoekstra, L., Harrison, S. J., Paxton, A., & Dixon, J. (2020a). Gesture–speech physics in fluent speech and rhythmic upper limb movements. *Annals of the New York Academy of Sciences*, 1491, 89–105.
- Pouw, W., & Fuchs, S. (2022). Origins of vocal-entangled gesture. *Neuroscience and Biobehavioral Reviews*, 141, 104836.
- Pouw, W., Harrison, S. J., Esteve-Gibert, N., & Dixon, J. (2020b). Energy flows in gesture-speech physics: The respiratory-vocal system and its coupling with hand gestures. *The Journal of the Acoustical Society of America*, 148, 1231.
- Pouw, W., Proksch, S., Drijvers, L., Gamba, M., Holler, J., Kello, C., Schaefer, R. S., & Wiggins, G. A. (2021). Multilevel rhythms in multimodal communication. *Philosophical Transactions of the Royal Society B*, 376, 20200334.
- Pouw, W., Werner, R., Burchardt, L., & Selen, L. (2024). The human voice aligns with whole-body kinetics. <https://www.biorxiv.org/content/10.1101/2023.11.28.568991v1>
- Prieto, P., Borràs-Comes, J., Tubau, S., & Espinal, M. T. (2013). Prosody and gesture constrain the interpretation of double negation. *Lingua*, 131, 136–150. <https://doi.org/10.1016/j.lingua.2013.02.008>
- Prieto, P., & Espinal, M. T. (2020). Prosody, gesture, and negation. In V. Depez & M. Teresa Espinal (Eds.), *The Oxford handbook of negation* (pp. 667–693). Oxford University Press.
- Rohrer, P. L., Tütüncübasi, U., Vilà-Giménez, I., Florit-Pons, J., Esteve-Gibert, N., Ren, A., Shattuck-Hufnagel, S., & Prieto, P. (2023). The multimodal multidimensional (M3D) labeling system. <https://doi.org/10.17605/OSF.IO/ANKDX>.
- Scott, S. V. & Orlikowski, W. J. (2014). Entanglements in practice: performing anonymity through social media. *MIS Quarterly*, 38 (3), 873–893.
- Thibault, P. J. (2011). First-order languaging dynamics and second-order language: The distributed language view. *Ecological Psychology*, 23(3), 210–245.
- Thibault, P. J. (2021). *Distributed languaging, affective dynamics, and the human ecology. Vol. 1: The sense-making body*. Routledge.
- Tubau, S., González-Fuente, S., Prieto, P., & Espinal, M. T. (2015). Prosody and gesture in the interpretation of yes-answers to negative yes/no-questions. *The Linguistic Review*, 32, 115–142. doi: 10.1515/trl-2014-0016
- Turk, A. E., & White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, 27(2), 171–206.
- Ward, N. G. (2019). *Prosodic patterns in English conversation*. Cambridge University Press.

---

**Cite this article:** Harrison, S. (2024). ‘This you may <sup>NNNN</sup>NEVER have heard before’: initial lengthening of accented negative items as vocal-entangled gestures, *Language and Cognition*, 1–34. <https://doi.org/10.1017/langcog.2024.26>