behavioral sciences

MDPI

Article

The Temporal Alignment of Speech-Accompanying Eyebrow Movement and Voice Pitch: A Study Based on Late Night Show Interviews

Volker Gast

Department of English and American Studies, Friedrich Schiller University, 07743 Jena, Germany; volker.gast@uni-jena.de

Abstract: Previous research has shown that eyebrow movement during speech exhibits a systematic relationship with intonation: brow raises tend to be aligned with pitch accents, typically preceding them. The present study approaches the question of temporal alignment between brow movement and intonation from a new angle. The study makes use of footage from the *Late Night Show with David Letterman*, processed with 3D facial landmark detection. Pitch is modeled as a sinusoidal function whose parameters are correlated with the maximum height of the eyebrows in a brow raise. The results confirm some previous findings on audiovisual prosody but lead to new insights as well. First, the shape of the pitch signal in a region of approx. 630 ms before the brow raise is not random and tends to display a specific shape. Second, while being less informative than the post-peak pitch, the pitch signal in the pre-peak region also exhibits correlations with the magnitude of the associated brow raises. Both of these results point to early preparatory action in the speech signal, calling into question the visual-precedes-acoustic assumption. The results are interpreted as supporting a unified view of gesture/speech co-production that regards both signals as manifestations of a single communicative act.

Keywords: eyebrows; facial gestures; pitch; prominence; audiovisual prosody

Citation: Gast, V. The Temporal Alignment of Speech-Accompanying Eyebrow Movement and Voice Pitch: A Study Based on Late Night Show Interviews. *Behav. Sci.* **2023**, *13*, 52. https://doi.org/10.3390/bs13010052

Academic Editor: Scott D. Lane

Received: 25 November 2022
Revised: 27 December 2022
Accepted: 30 December 2022
Published: 6 January 2023

1. Introduction

1.1. Context of the Study: Multimodal Communication

Human face-to-face communication is multimodal, primarily relying on acoustic and visual signals [1]. The acoustic signal is produced in the vocal tract and is constituted by the segments of speech, such as [w], [ɛ], and [l] in the word *well*. The speech segments are realized with a specific prosody, i.e., melody and rhythm. For example, the word *well* can be pronounced with varying degrees of length, and different types of tone movement, e.g., falling ("Well!") or rising ("Well?"). Prosody has various meta-propositional functions, such as the expression of emotions and attitudes [2], the indication of prominence [3], and the organisation of the conversational exchange (e.g., turn-taking) [4].

The visual signal is produced with visible body parts, most prominently the hands and the face, in the form of gestures [5,6]. There is comprehensive literature on hand gestures in the tradition established by Adam Kendon and David McNeill [7,8], covering aspects such as their structure ('preparation', 'stroke', 'retraction') [9–11], their degree of 'autonomy' relative to speech (from speech-accompanying gestures to the signs of sign languages [12–14]), and their semiotic properties ('iconic', 'metaphoric', 'deictic', 'beat' [13]). Research on facial expressions has an even longer tradition, starting with the work of Paul Ekman and Wallace Friesen in the 1960s. Ekman and Friesen investigated both the display of emotions [15–17] and the communicative signals [18–20] associated with facial expressions.

Even though there was some cross-fertilization between the study of hand gestures and facial expressions at an early stage, e.g., in the work by A. Kendon [9,10,21] and D.

McNeill [13,22], it was only in the 1990s that the two types of signals started to be analyzed in tandem in a more systematic way. For example, the *Integrated Message Model of Language* proposed by Janet Bavelas, Nicole Chovil and colleagues analyze 'conversational facial gestures' with the toolbox developed for hand gestures [23–29], intending "to articulate and document the extensive similarities of facial to hand gestures, which offer an alternative to approaches that see the face as stereotypic configurations related to a few emotional expressions" [28] (p. 31).

A central, still open question of research into multimodal communication concerns the relationship between speech and visual signals from the point of view of production. The two types of signals are produced concurrently, and the question arises of whether they "share a computational stage", in McNeill's words [22] (p. 350)—and if so, what stage(s) they share. Put differently, the question is whether gesture and speech are two manifestations of the same action, or whether they are separate actions (see [30] for an overview of the discussion). According to the 'separatist' view, as expressed, for instance, in the 'Lexical Retrieval Theory' [31–34], speech and gestures are processed separately, and gestures have an auxiliary role relative to, and facilitative effect on, speech. In fact, proponents of this view hold that gestures contain very little information to begin with, and that "the gestural contribution to communication is, on the whole, negligible" [33] (p. 93). According to the unified view, often associated with McNeill's 'Growth Point Theory' [22], "the speech production process is manifested in two forms of activity simultaneously: in the vocal organs and also in bodily movement" [10] (p. 211). Gestures and speech are taken to emerge from the same underlying conceptualizations, reflecting different facets of the same message [35–39].

One type of evidence that has been used in support of a unified view of speech-gesture co-production concerns the temporal coordination of speech and gestures. Goldin-Meadow and Brentari [40] write:

> gesture and speech are temporally organized as a single system. The prosodic organization of speech and the phrasal structure of the co-occurring gestures are coordinated so that they appear to both be produced under the guidance of a unified plan or program of action [...]. For example, the gesture and the linguistic segment representing the same information as that gesture are aligned temporally. More specifically, the gesture movement—the "stroke"—lines up in time with the tonic syllable of the word with which it is semantically linked (if there is one in the sentence). [40] (p. 16)

Advocates of a separatist view, however, point to the well-known asynchrony of gestures and speech as evidence for an early split in the production of gestures and speech. Gestures generally tend to precede the segment in the speech signal that they relate to (the 'lexical affiliate') [33,41–43]: "[t]he idea that lexical gestures have an early origin is consistent with the well-established finding that lexical gestures precede their lexical affiliates ...." [33] (p. 23).

While there is broad consensus that gestures and speech are coordinated but asynchronous, with gestures normally preceding their lexical affiliate, the details of crossmodal temporal coordination are not yet fully understood. Studying that coordination is therefore a desideratum of research into multimodal communication: "it appears that a fully comprehensive model of the speech planning process must include a mechanism that can provide an account of speech-gesture alignment in time and the relationship of both speech and gesture to structure and meaning" [44] (p. 1210). The present study intends to contribute to this agenda by investigating the temporal coordination of a prominent facial articulator, the eyebrows, and a central aspect of prosody, i.e., intonation.

1.2. Facial Gestures and Audiovisual Prosody

Eyebrows are associated with specific communicative functions. Four types of functions have figured prominently in research on eyebrow gestures: (i) the expression of

emotions, (ii) the expression of epistemic attitudes, (iii) the expression of specific types of illocutionary acts, and (iv) the handling of turn-taking in conversation. The most prominent emotion linked to eyebrow movement is that of surprise. This association has often been noted [45–48] and is, in a way, archetypical from an evolutionary point of view. Eyebrows are raised as a result of opening one's eyes wide, as a symptom of "attentional activity" [20,48–50]. Related to the expression of surprise is the epistemic attitude of uncertainty or ignorance. Ignorance is encoded in a conventionalized gesture, the 'facial shrug' [27,29]. Moreover, it has been shown that raised eyebrows tend to accompany modals like *may* and *might* [51]. As indicators of illocutionary force brow raises have been claimed to signal yes-no question [20,52–54], but this claim has also been called into question [51,55,56]. Flecha-García [55,56] found that brows tend to be raised in 'instruct moves' in dialogue games used for elicitation (the 'Map task' [57,58], e.g., "So then you go along just a few dashes" [55] (p. 68)). The role of eyebrow movement in turn-taking has been noticed in several studies [55,59–61], e.g., insofar as, "more frequent and longer eyebrow-raising occurred in the initial utterance of high-level discourse segments than anywhere else in the dialogue" [55] (p. ii).

Beyond the expression of communicative functions of the type mentioned above, eyebrows have been found to signal the 'prominence' of their affiliates in a prosodic sense. A parallelism of gestures and prosody has long been noted from the perspective of prosody [62]. D. Bolinger holds that "[t]he fluctuations of pitch are to be counted among all those bodily movements which are more or less automatic concomitants of our states and feelings and from which we can deduce the states and feelings of others. Intonation belongs more with a gesture than with grammar" [63] (p. 157). Just like the pitch trajectory moves up and down, "[o]ther up-down gestures can be carried by the eyebrows, the corners of the mouth, the arms and hands, and the shoulders. Motion in parallel with pitch is again the rule" [63] (p. 158). The assumption is that prosody and gestures cooperate in conveying prominence can be termed the *Audiovisual Prosody Hypothesis*.

Several studies have provided evidence for the Audiovisual Prosody Hypothesis using eyebrow data. Cavé et al. [64] studied the relationship between brow raises and pitch movement in elicited data. The authors manually annotated their data for brow raises, and determined the type of pitch movement at the time of the raise (rising-falling, falling-rising, plateau, rising and falling). They found an association of brow raises with contours containing a rise (rising-falling, falling-rising, and rising). These results were confirmed in a later study [59].

Flecha-García investigated (among other things) the temporal distance between brow raises and pitch accents in dialogic recordings of four subjects performing the Map Task [57,58]. With manual annotations she identified both brow raises and pitch accents, and determined distances between them. She found that the distance between brow raises and the closest pitch accent is close to zero (μ = 63 ms, σ = 458 ms), providing evidence of alignment. Moreover, brow raises "are significantly closer to their following PA [pitch accent] than to their preceding PA" [55] (p. 126). This shows that brow raises, like hand gestures, tend to precede the segment in the speech signal that they relate to.

Kim et al. [54] used elicited data from six participants to study the association of eyebrow and head movement in relation to different types of information-structural configurations (broad focus, narrow focus, echoic question). The narrow focus and echoic question conditions were associated with eyebrow action in a region of ≈250 ms preceding the prosodically prominent constituent. The tendency of brow raises to occur in the neighborhood of prominent words, and to precede those words, was thus confirmed.

Following up on Kim et al. [54], Schade and Schrumpf [65] investigated the alignment of brow raises and pitch accents, taking the additional dimension of 'emotion' into account. Their hypothesis was that "f0 peaks and eyebrow movement peaks [would] be temporally closer in emotional utterances" [65] (p. 95). The authors used elicited data—story retellings from an egocentric perspective—annotated with OpenFace [66] for eyebrow movement, and with Praat [67] for the audio signal. The study confirmed the results

obtained by Kim et al. [54]—"pitch and eyebrow movement are temporally aligned" and showed moreover that "[e]motionality seems to enhance this alignment" [65] (p. 99).

Unlike the other studies mentioned above, Berger and Zellers [68] used naturalistic data—semi-spontaneous YouTube monologs—to investigate the interaction of intonation and eyebrow movement. On the basis of automatic OpenFace annotations and manual pitch annotations, the authors found evidence of correlations between "pitch height and eyebrow movements . . . for at least some of the measures for all speakers" [68] (p. 1). Moreover, they went beyond previous studies aiming to "investigate the relationships between data in the form of contours rather than individual points" [68] (p. 10). For each pitch accent, they determined eyebrow trajectories around the F0-peak in a window of 1 s. Using Functional Principal Component Analysis, the authors did not find evidence for synchronization of eyebrow peaks and F0 peaks, contradicting the findings of Flecha-García [55]. However, it is important to keep in mind that their method differed from that of Flecha-García: while Flecha-García took brow raises as a point of reference and identified pitch accents in their neighborhood, Berger and Zellers used pitch accents as a point of departure and identified brow raises in their neighborhood. The authors also point out a few caveats, e.g., that the eyebrow positions were generally rather high in the 1-s window around the pitch accents, and that cases of brows in rest position were excluded.

The present study follows up on Berger and Zellers [68] in two respects. First, it uses naturalistic, rather than elicited data. And second, it investigates contours, in addition to point measurements. Pitch is mostly studied in one of two ways, in the field of phonology. It can either be regarded as being constituted by pitch trajectories (rise, falls, rise-falls . . .), or by pitch levels or targets (high, low). The first approach is typically associated with what is often called the 'British school' of intonation [69]; the second one, particularly the 'autosegmental' paradigm, has its roots in the US and is therefore often called the 'American school' [70]. The present study takes the perspective of the British school of intonation and thus treats pitch as a dynamic, rather than positional, signal. This has consequences for the question of temporal alignment between eyebrow positions and pitch. Eyebrow positions can reasonably be assumed to constitute a positional signal: a high position is associated with a high degree of prominence. This is different for pitch: prominence is not associated with a particularly high fundamental frequency, but with a high degree of change in the signal. From this point of view, we would wish to correlate eyebrow positions with properties of pitch contours, rather than point measurements.

There are important differences between the present study and the one by Berger and Zellers. First, the data is dialogic, as in most of the other previous studies mentioned above. As Berger and Zellers [68] note, eyebrow movement in YouTube monologs may not be entirely natural. Second, the study is completely data-driven, using no manual annotations, and thus not relying on any theoretical preconceptions (which may be controversial in the domain of phonology). Unlike Berger and Zellers [68], and like Flecha-García [55], the brow raises are moreover taken as a point of reference, and pitch movement is investigated in relation to brow raises. The reason is that—as pointed out above—eyebrow height is a (positional) point measurement, with the highest measurement corresponding to the highest visual prominence level, whereas pitch contours are regarded as dynamic signals whose prominence is reflected in degrees of change, not absolute values. Finally, the present study differs from that of Berger and Zellers [68] in that it refrains from using high-level, and potentially black-box, methods of analysis, remaining close to the 'observable' signal. Pitch contours are treated as near sinusoidal functions. The parameters of these functions—specifically the amplitude and the ordinary frequency—are subject to constant change, but relatively stable within individual cycles (up-and-down movements). The parameters of the sinusoidal functions fitted to the data are regarded as correlates of 'dynamicity', and are used as input to the quantitative analysis.

1.3. Objectives, Significance and Structure of the Study

The present study is intended as a contribution to our understanding of the relationship and interaction between prosody in the speech signal and eyebrow movement as a prominent facial articulator. Based on the Audiovisual Prosody Hypothesis, which holds that acoustic and visual signals jointly convey prominence, it studies pitch movement in the neighborhood of brow raises. Importantly, the pitch is not only treated as a series of point measurements, but also as a sinusoidal function. This approach allows us to determine not only levels of pitch, but also degrees of dynamicity (amplitude and frequency of the signal), and the shape of the signal.

The results confirm observations made in previous studies, specifically concerning the temporal alignment of brow raises and pitch. The data show that brow raises tend to be followed by higher-than-average pitch values. Moreover, the post-peak region shows more fluctuations in the signal, i.e., higher degrees of dynamicity.

The study also arrives at new results. The shape of the pitch contour in the pre-peak region correlates with the height of the brow raise. In a window of approx. 630 ms before the brow raise, a later start in the phase of a sine wave–with the pitch falling, or rising from a low level–correlates with higher brow peaks. This finding has potential implications for the co-production of brow raises and prosody. While it is generally assumed that the visual signal (the brow raise) tends to precede the acoustic signal (the pitch accent), the shape of the pitch contour in the pre-peak region points to a 'preparatory' phase before the actual prominence marking, and can be regarded as an early symptom of prominence. The observation of a substantial preparatory phase in the pitch contour sheds doubt on the visual-precedes-acoustic assumption.

The study is structured as follows: Section 2 contains a description of the data and the methods of analysis. Section 3 presents the results, which are discussed in Section 4.

2. Materials and Methods

2.1. The Corpus and the Processing of the Data

The present study is based on footage from the *Late Night Show with David Letterman* recorded between 1980 and 2014. The material was obtained from a fan page (https://donzblog.home.blog/, accessed on 29 December 2022). The clips come with automatically generated subtitles (in srt-format), with a reasonable, though not perfect, quality. The corpus comprises 160 files, with some files covering several shows. Altogether, the corpus contains approx. 160 h of recording. Obviously, not all of the material can be used for the analysis of facial gestures in relation to prosody, as the camera angles vary, there is overlapping speech (introducing noise into the audio signal) and sometimes, the camera focuses on the guest when the host speaks. A sample of video sequences was therefore created manually, as detailed below.

In the first step, faces were recognized in the video signal, using the dlib-package for Python (https://pypi.org/project/dlib/, accessed on 29 December 2022). Only sequences of frames were used that contained at least three seconds of recording of the same face, from the same angle, at a resolution of at least 10,000 pixels per face. The video data was processed at a rate of 30 frames per second. All measurements are associated with frames, and 30 subsequent measurements correspond to one second of recording. The faces thus recognized were annotated for facial landmarks using 3D facial landmark detection software [71] (https://github.com/1adrianb/face-alignment, accessed on 29 December 2022). Thefacial landmarks delivered by this package are comparable to those of OpenFace (https://github.com/TadasBaltrusaitis/OpenFace, accessed on 29 December 2022), though no Action Units were determined, in keeping with the data-driven approach of the study. The software identifies 68 facial landmarks, which can be used to measure out the face of the speakers, and to create a model of the speaker's face, see Figure 1 on the left for an example.

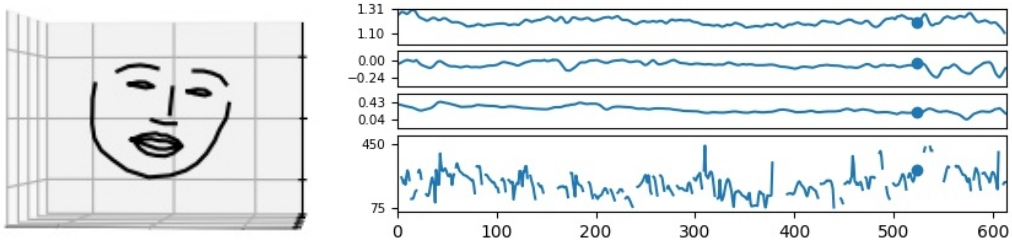

Figure 1. A frame with 3D facial landmark annotation (left) and measurements of eyebrow height, sagittal head angle, lateral head angle, and pitch; the segment has been taken from Episode 0148 featuring Art Donovan. Original frames are not shown for reasons of copyright.

The facial landmark annotations were used to obtain measurements for three variables relevant to the study of facial gestures: (i) the height of the eyebrows, (ii) the lateral angle of the head, and (iii) the sagittal angle of the head (the calculation of the angles is only possible with 3D models). The height of the eyebrows was measured as the ratio of the distance between the centroid of the four landmarks delimiting the nose, and the centroid of the ten points marking the eyebrows. The measurements were standardized by a speaker.

The acoustic signal was processed with Praat [67], at a rate of 30 measurements per second, like the video data. The data thus obtained was visualized for manual inspection as shown on the right side of Figure 1. The 3D-model of the speaker is shown on the left, and the four signals of interest are displayed on the right (from top to bottom, eyebrow height, lateral head angle, sagittal head angle, and pitch; the measurements are contained in the file df_measurements.tsv in the Supplementary Materials).

2.2. The Sample Used for the Analysis

The corpus was partitioned into *segments* corresponding to subtitle units as represented in the automatically generated subtitle files. This way of partitioning the data allowed us to control the lexical material associated with the pitch and eyebrow data, which proved particularly useful when it came to manually inspecting and searching the corpus. We regard the segments as a random sample of sequences from the corpus, with mean durations of 2.28 s ($\sigma = 1.08$). A sequence of such segments is shown in (1) and can be inspected by clicking 📷. The segments are identified by the number of the 'episode' (mp4-file) and the subtitling segment (e.g., 0145-2372 for segment 2372 in the subtitle file 0145.srt)

(1) 📷

a. …this movie and the same guy did my [0145-2372]
b. makeup for coming to America Rick Baker [0145-2373]
c. he turned me into this 400 yeah he [0145-2374]
d. turned me to this 400-pound guy and uh [0145-2375]

To control for speaker-specific effects, a sub-sample was created—the *speaker sample* in the following—with data from five speakers who figure prominently in the show (as recurrent guests), and whose facial landmarks did not show any systematic errors: Art Donovan (AD), Buck Henry (BH), Eddie Murphy (EM), Teri Garr (TG), and Quentin Crisp (QC). The sample was manually inspected to make sure that the interviewee, who is filmed, speaks while being filmed. The sample thus created consists of 887 segments, with 55,504 frames and 1850 s (AD: 339 segments, 20,552 frames, 685.1 s; BH: 97 segments, 6972 frames, 232.4 s; EM: 151 segments, 8385 frames, 279.5 s; QC: 11 segments, 9180 frames, 306.0 s; TG: 187 segments, 10,415 frames, 9180 frames, 347.2 s; the measurements for the sample are contained in the file df_sample.tsv in the Supplementary Materials).

For the study of pitch contours in the environment of brow raises the data was processed as follows: we identified in each segment the position where eyebrow height was at its maximum, the 'eyebrow peak'. 120 measurements around the eyebrow peak were extracted, up to 60 pitch measurements preceding it and up to 60 measurements

following it. This gave us, for each corpus segment, a series of 121 measurements (as the pitch and eyebrow height at the peaks were also included), and a window of 2s around the eyebrow peak (the pitch measurements around the eyebrow peak are contained in the file df_sample_pitch.tsv, and the corresponding eyebrow measurements are contained in the file df_sample_eb.tsv in the Supplementary Materials).

Figure 2 shows the series of eyebrow measurements for the speaker sample. The raises exhibit comparable widths, covering approx. a period of ten to twenty frames around the peak (≈300–600 ms). The raises seem to be largely symmetrical. A plot showing the eyebrow series for the entire dataset is provided in the Appendix A (Figure A1). According to a Mann-Whitney U test, the eyebrow heights between positions −8 and 9 (in a window of approx. 270–300 ms around the peak) are significantly higher than the mean. I assume that this window approximately corresponds to the scope of the brow raise.

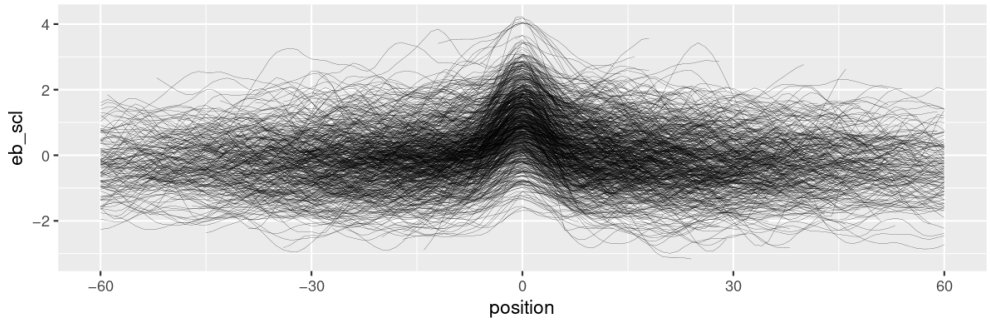

Figure 2. Time series for eyebrow movement around the peak of a segment (speaker sample). The values are scaled per speaker.

2.3. Fitting Sinusoidal Models to the Data

To detect correlations between eyebrow height and pitch measurements, sinusoidal models were fitted to windows of different lengths to the data preceding the eyebrow peak (the 'pre-peak region') and the data following it (the 'post-peak region'), filling in missing values with linear interpolation. While pitch does not generally follow a sinusoidal pattern, the individual ups and downs characterizing the tones of English can be captured with such a model. A sine wave is standardly described in terms of three parameters: (i) the amplitude (A), (ii) the ordinary frequency (f), (iii) and the phase (ϕ). The amplitude corresponds to the difference between the center of the wave and the absolute values of its highest and lowest points. The ordinary frequency represents the number of cycles per time unit, and thus the rate of pitch changes. The phase ϕ, measured in radians, corresponds to the part of the cycle where the waveform starts, and ranges from 0 to 2. I will refer to the phases as 'high-rising' ($0 \leq \phi < 0.5$), 'high-falling' ($0.5 \leq \phi < 1$), 'low-falling' ($1 \leq \phi < 1.5$) and 'low-rising' ($1.5 \leq \phi < 2$). Figure 3 illustrates a sine wave with the ϕ-value (for the phase) on the x-axis.

In addition to the standard parameters of a sine wave we included a y-intercept, i.e., a value shifting the entire curve up or down on the y-axis. This intercept (I) will be called the 'baseline shift'. (2) shows the model equation for the sine wave (t is a point in time). The models were fitted using the function `nls()` of R [72] (the parameters for the best-fitting curves are contained in the files res_df_nls_pre.tsv for the pre-peak region, and res_df_nls_post.tsv for the post-peak region.)

$$(2) \quad y = I + A \times sin(2 \times \pi \times f \times t + \phi \times \pi)$$

The sinusoidal models fitted to the pre- and post-peak regions of two segments are shown in Figure 4, for a window size of 17 frames (≈570 ms). The blue lines show the pitch measurements, the green lines represent the sinusoidal model fitted to the 16 measurements preceding the eyebrow peak as well as the peak, and the red line shows the model fitted to the pitch contour following the peak. The black line shows the value

of the eyebrow peak (in standard deviations per speaker). The post-peak model in the upper plot shows a sine wave with an amplitude of 3.3, an ordinary frequency of 0.37, and a ϕ-value of 1.04 (i.e., it starts in the early low-falling phase). The post-peak model in the bottom plot shows a sine wave with an amplitude of 2.4, an ordinary frequency of 0.86, and a ϕ-value of 0.24 (it starts in the late high-rising phase). The contour in the bottom plot is more dynamic than the one in the upper plot insofar as both the range (A) and rate (f) of pitch values is higher.

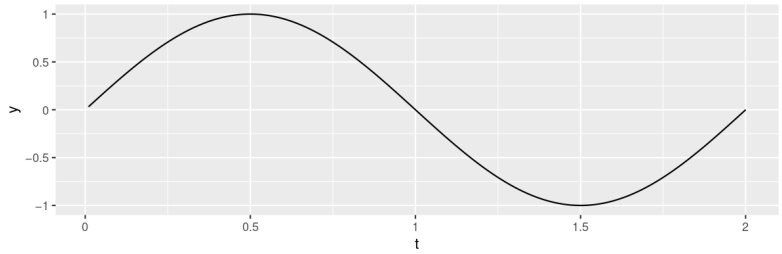

Figure 3. Sine wave, with phases on x-axis.

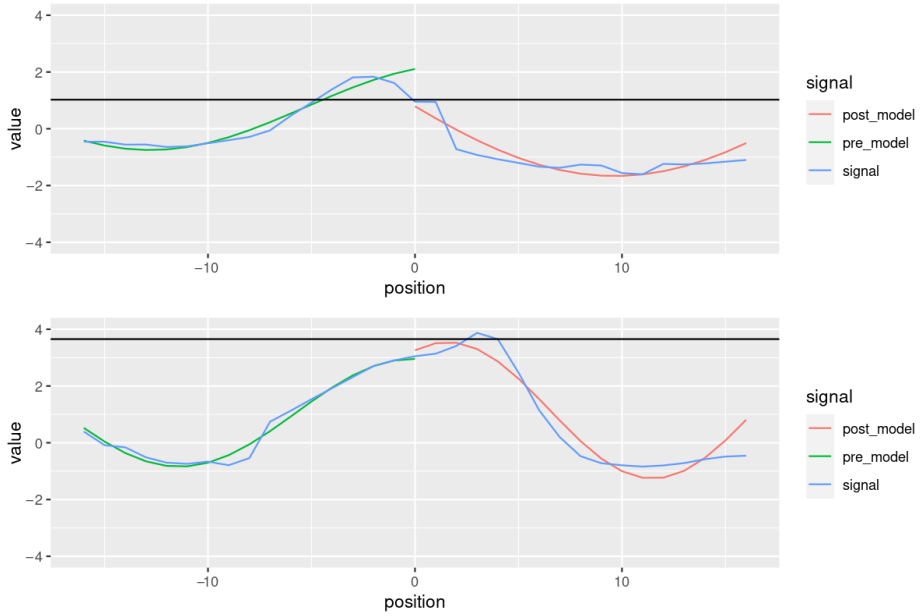

Figure 4. Two pitch contours with sinusoidal fittings (segments 0099-0249 and 0099-0337, both by QC). The blue line shows the pitch contour, the green line shows the sine curve fitted to the pre-peak region, and the red line shows the sine curve fitted to the post-peak region.

In the statistical analysis, the parameters determined for each pre- and post-peak contour were used as predictors for the eyebrow height at its peak, using mixed regression modeling, with 'speaker' as a random effect (intercept only).

3. Results

3.1. Correlations between Pitch and Eyebrow Height Measurements

Figure 5 shows a smoothed curve (generalized additive model, k = 30) representing the eyebrow height (blue line) and pitch values (both standardized per speaker) with 95% confidence intervals for the standard error. The pattern is remarkably clear. Eyebrow movement is symmetrical (see also Figure 2), while the pitch curve drops before the eyebrow peak, then raises, crossing the zero line approximately at the peak, and remains at a higher-than-average level for a few frames.

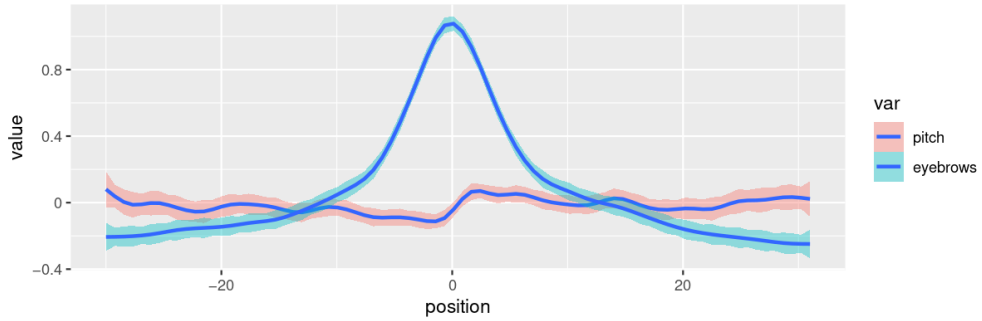

Figure 5. Smoothed curves showing the eyebrow height (blue) and pitch value (red) of the speaker sample, with 95% confidence intervals for the standard errors. The smoothing curves show a generalized additive model fitted to the data (k = 30).

Figure 5 aggregates all types of brow raises—very low ones, which are likely not intended as facial gestures but purely co-articulatory, and high ones, which (may) convey a meaning of their own. Figure 6 shows the curves for eyebrow peaks with a height of <1.5 standard deviations (top plot), and those with a peak of >1.5 standard deviations (bottom plot). The low raises show a pattern where the pitch is well beneath average and rises to reach an average height shortly after the eyebrow peak. In the neighborhood of the higher raises, the pitch is at an average level before the eyebrow peak and raises from there to a higher-than-average value for a certain period. The generalized additive model (with k = 10) shows the lower bound of the standard error to be >0 for positions −1 to 19. This confirms the results obtained in studies showing that pitch tends to be elevated in the region following a (gestural) brow raise.

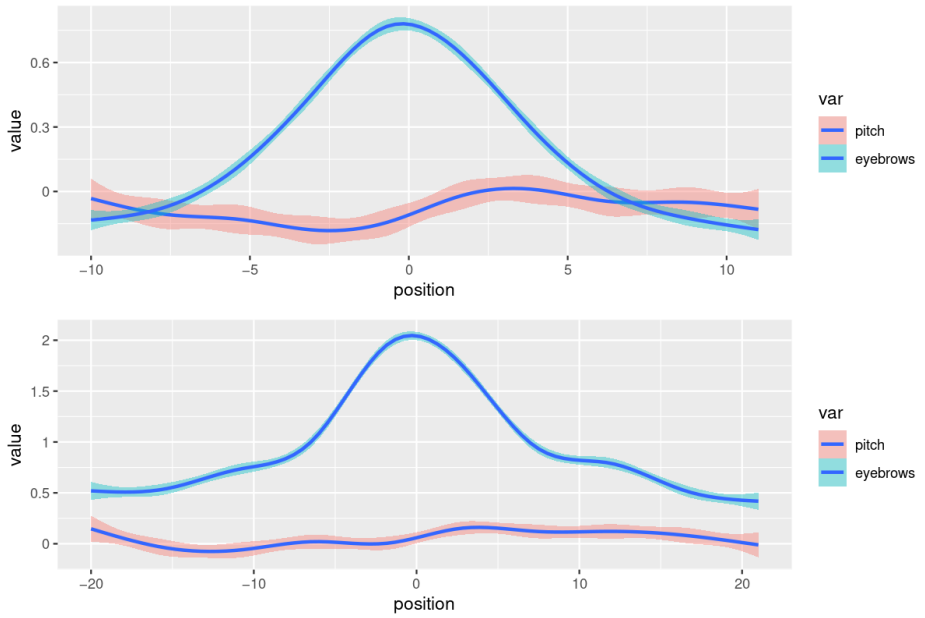

Figure 6. Smoothed curves showing the eyebrow height (blue) and pitch value (red) of the speaker sample, with 95% confidence intervals for the standard errors, for segments with an eyebrow peak of <1.5 standard deviations (top plot) and an eyebrow peak of >1.5 standard deviations (bottom plot). The smoothing curves show generalized additive models fitted to the data (k = 10).

3.2. Correlations between Eyebrow Measurements and the Parameters of Sinusoidal Functions

The sinusoidal models fitted to the pitch contours generally show a rather good fit, with the coefficients of determination averaging at $r^2 = 0.58$, for both the pre- and post-peak regions, for a window size of 20. Histograms for the r^2-values of these models are shown in

Figure A2 in the Appendix A. R^2-values do not show any correlations with other variables used for the analysis, i.e., the goodness of fit does not vary with the type of contour.

Mixed-effects models (with a random intercept for 'speaker') were fitted to pitch series of different window sizes, from 10 to 20 frames. The parameters of the sinusoidal functions fitted to the pre-peak region show significant correlations in the larger range of window sizes. The best model was obtained at a window size of 19 frames.

Table 1 shows the estimates for the main effects of this model, with the standard errors and the results of a t-test, as delivered by the `lmer()`-function of the `lmerTest`-package for R [73]. The parameters for amplitude (A) and ordinary frequency (f) show a positive correlation with the value of the eyebrow height at its peak. Moreover, the phase parameter ϕ shows a positive correlation with the maximum eyebrow height. This means that the contours accompanying higher brow raises tend to start in the later phases of a sine wave. There is no correlation between the maximum eyebrow height and the baseline shift. No interactions between the parameters were found.

Table 1. Estimates for main effects of a mixed linear model predicting the magnitude of the eyebrow peak in a segment based on the parameters of the sinusoidal function for the pre-peak region, with a window size of 19 frames (≈630 ms).

	Estimate	Std. Error	df	t Value	Pr(>|t|)
Intercept	0.645	0.122	82.693	5.270	<0.001
A_{pre}	0.043	0.021	609.598	2.027	0.043
f_{pre}	0.240	0.090	609.332	2.675	0.008
ϕ_{pre}	0.197	0.078	609.991	2.538	0.011

For illustration, consider Figure 7, which shows sinusoidal models fitted to the pre-peak region at a window size of 19 frames that start with a high-falling (left) and a low-falling contour (right). The plot only shows pitch trajectories of segments with an eyebrow peak of >1.5. It is this type of contour that tends to accompany higher brow peaks.

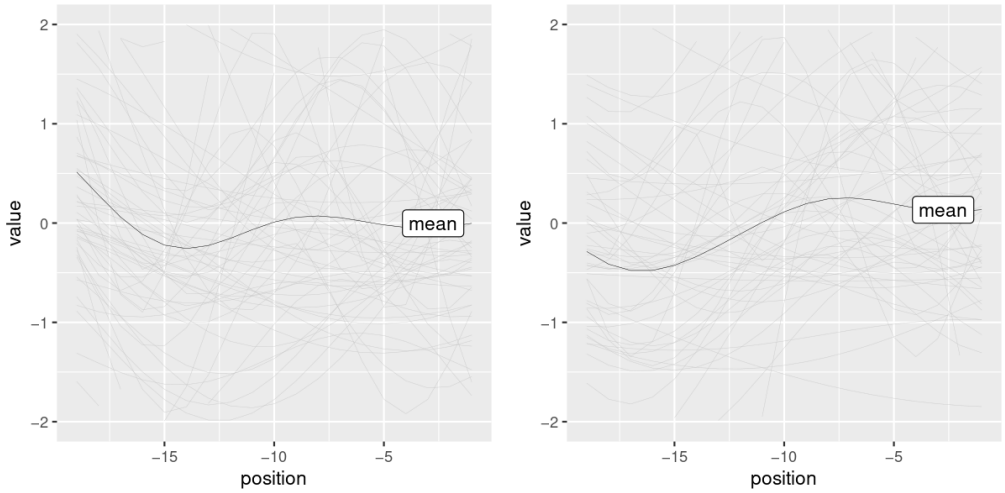

Figure 7. Fitted sinusoidal curves for the pre-peak regions, within a window of 19 frames. The thick lines in the center show the mean values.

The best model fitted to the post-peak region was based on the parameters of models fitted to a window of 17 frames. The linear model shows similar—in fact, stronger—effects for amplitude (A) and ordinary frequency (f) than the pre-peak model, but the phase has no significant effect. As in the case of the pre-peak region, there is no correlation with the baseline shift. Again, no interactions were found. The estimates of a model with significant predictors only are shown in Table 2.

Table 2. Estimates for main effects of a mixed linear model predicting the magnitude of the eyebrow peak based on the parameters of the sinusoidal function for the post-peak region, with a window size of 17 frames (≈570 ms).

	Estimate	Std. Error	df	t Value	Pr(>|t|)
Intercept	0.633	0.120	38.261	5.254	<0.001
A_{post}	0.153	0.049	658.456	3.149	0.002
f_{post}	0.349	0.090	658.474	3.885	<0.001

To compare the correlations between the pre- and post-peak phase with the height of the eyebrow peak, we fitted a model with variables from both regions, with the parameters for the window sizes providing the best models for either region (19 frames for the pre-peak region and 17 frames for the post-peak region). The model includes all those variables that have a significant effect in the separate models for the pre- and post-peak regions. Table 3 shows the estimates.

Table 3. Estimates for main effects of a mixed linear model predicting the magnitude of the eyebrow peak based on the parameters of the sinusoidal function for the pre-peak region, with a window size of 19 frames (≈630 ms), as well as the post-peak region, with a window size of 17 frames (≈570 ms).

	Estimate	Std. Error	df	t Value	Pr(>|t|)
Intercept	0.213	0.210	210.104	1.012	0.313
A_{pre}	0.038	0.065	397.280	0.584	0.560
f_{pre}	0.217	0.120	396.176	1.817	0.070
A_{post}	0.162	0.059	396.820	2.738	0.006
f_{post}	0.387	0.110	397.808	3.523	<0.001
ϕ_{pre}	0.238	0.101	397.285	2.360	0.019

In a combined model, the amplitude of the pre-region does not show a significant effect (p = 0.56). The ordinary frequency of the pre-peak region exhibits a tendency, but is not significant at α = 0.05 (p = 0.07). The ϕ-value of the pre-peak region remains significant (p = 0.02). The effects for amplitude and ordinary frequency of the post-peak region show similar magnitudes and levels of significance as in the model fitted to the post-region only.

4. Discussion and Conclusions

Previous studies have shown that brow raises typically (though not necessarily) occur in the vicinity of pitch accents, either synchronous with or preceding them. This tendency has been confirmed by the data analyzed in the present study. Pitch values tend to be higher than average in a region of approx. 670 ms following the peak of the brow raise in those cases where the eyebrow peak has a height exceeding 1.5 standard deviations. This seems to show that the visual signal (the brow raises) does in fact precede the acoustic signal (pitch).

The results obtained on the basis of absolute pitch measurements have to be interpreted with a grain of salt, however. It seems reasonable to use positional measurements for the height of the eyebrows, as operationalizations of visual prominence: higher brows convey more prominence than lower brows. By contrast, in the acoustic signal prominence is conveyed through the degree and rate of change, not through absolute position. This means that to understand the temporal alignment between the two signals, we need to compare degrees and rates of change in the acoustic signal with absolute values in the visual signal.

In the present study the dynamic nature of the pitch signal was taken into account by not comparing positional values of eyebrow positions with absolute pitch measurements, but with properties of the pitch contours as reflected in the parameters of sinusoidal models fitted to the data. In this way we obtained measurements of the degree of 'dynamicity' of a contour (amplitude/A and ordinary frequency/f), and its shape (as reflected in the phase ϕ).

The results show that the amplitudes and ordinary frequencies in the post-peak region exhibit a positive correlation with the maximum height of the eyebrows in a brow raise. No preference for a specific shape could be identified in this region. A model fitted to the pre-peak region shows similar effects as far as dynamicity is concerned. However, the effects of amplitude and ordinary frequency disappear in a combined model with the post-peak region. The pre-peak values for A and f are less informative than the post-peak values, as reflected in their lower regression coefficients, and they do not provide additional information in comparison to the post-peak values. Still, it should be borne in mind that when considered on their own, they do show significant correlations with the eyebrow height at its peak.

Perhaps the most interesting finding about the pre-peak region is that the shape of the signal is informative. Within a window of 19 frames preceding the peak, there is a significant tendency for the contours to start at a later stage in the cycle of a sine wave. This means that the pitch is either falling, or rising from a low position. The fact that the shape of the pitch contour at that stage is not random points to a preparatory phase in the expression of prominence. The exact contours need to be investigated further, with attention to more potential predictors, such as the lexical material in the context of the brow raise.

As for the more general question concerning the relationship between visual and acoustic signals, the present study confirms the 'visual-first' tendency observed in earlier studies. However, there are 'early symptoms' of prominence—and, hence, temporal alignment—as well. The period at which the shape of the contour is informative, relative to the position of the eyebrows at the peak, spans approx. 630 ms and thus precedes the onset of the brow raise, which (typically) starts approx. 270 ms before the peak.

The early symptoms of prominence in the acoustic signal may indicate that the pitch contour has a relatively long planning phase, in comparison to the eyebrow movement. The relative lateness of the perceptible symptoms of prominence in the acoustic signal could thus be due to the fact that pitch cannot be changed instantly, as it constitutes a time series with auto-correlations. Pitch forms structured, near-sinusoidal signals with regular and rule-governed ups and downs, and prominence marking needs to be integrated into that signal, which requires adaptations spanning a certain period of time, potentially leading to a lag in comparison to the visual signal.

With respect to the question of the "computational stages" shared by gestures and speech [22], the results of the present study are compatible with a one-action view of gesture/speech co-production. In particular, the degree of asynchronicity of the two signals has been relativized. The early symptoms of prominence in the acoustic signal point to a preparatory phase of ≈270 ms preceding the brow raise. While the two signals are not perfectly aligned, they seem to be co-generated in a systematic and structured way, following a "unified plan or program of action" [40] (p. 16).

Supplementary Materials: The Supplementary Material with the data can be downloaded at: https://doi.org/10.5281/zenodo.7485020.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original footage is copyright protected but can be inspected at https://donzblog.home.blog/, accessed on 29 December 2022. The relevant measurements are made available in the Supplementary Materials.

Acknowledgments: I wish to thank the participants of the Olomouc Linguistics Colloquium (Olinco) 2020, and of the Workshop on 'Visual Communication. New Theoretical and Empirical Domains' (within the Annual Conference of the Deutsche Gesellschaft für Sprachwissenschaft) 2021 for valuable comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

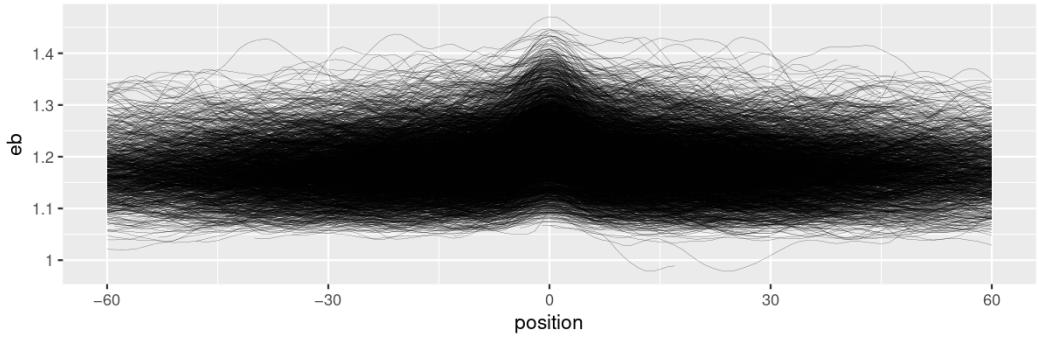

Figure A1. Time series for eyebrow movement around the peak of a segment (complete sample). The values are scaled per segment.

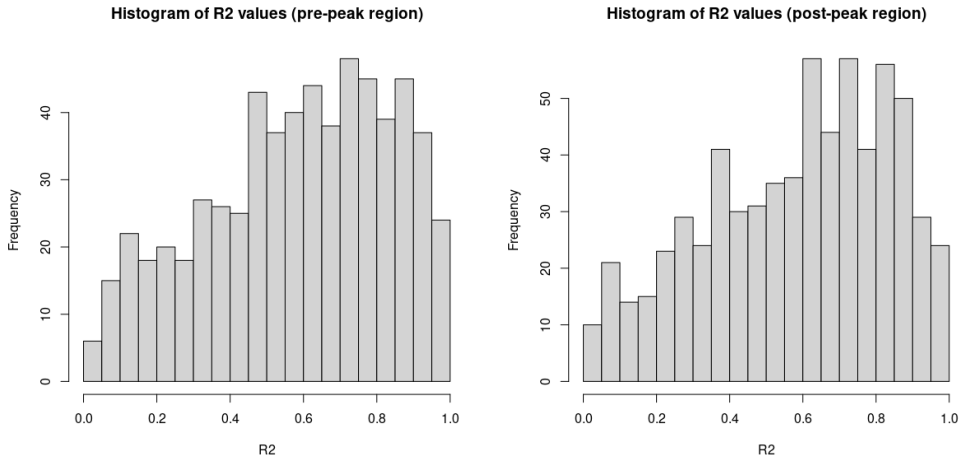

Figure A2. Histograms for the r-squared values of the models fitted to the pre-peak region (**left**) and the post-peak region (**right**), with a window size of 20 frames.

References

1. Allwood, J.; Ahlsén, E. Multimodal communication. In *Handbook of Technical Communication*; Mehler, A., Romary, L., Eds.; de Gruyter Mouton: Berlin, Germany, 2012; pp. 435–460.
2. Wichmann, A. The attitudinal effects of prosody, and how they relatet to emotion. In Proceedings of the ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion, Newcastle, UK, 5–7 September 2000; Cowie, R., Douglas-Cowie, E., Schröder, M., Eds.; 2000; pp. 143–148.
3. Grice, M.; Kügler, F. Prosodic Prominence—A Cross-Linguistic Perspective. *Lang. Speech* **2021**, *64*, 253–260. [CrossRef] [PubMed]
4. Cutler, A.; Pearson, M. On the analysis of prosodic turn-taking cues. In *Intonation in Discourse*; Johns-Lewis, C., Ed.; Routledge: London, UK, 1986; pp. 139–155.
5. Müller, C.; Cienki, A.; Fricke, E.; Ladewig, S.H.; McNeill, D.; Bressem, J. (Eds.) *Body—Language—Communication. An International Handbook on Multimodality and Human Interaction*; De Gruyter Mouton: Berlin, Germany, 2013; Volume 1.
6. Müller, C.; Cienki, A.; Fricke, E.; Ladewig, S.H.; McNeill, D.; Bressem, J. (Eds.) *Body—Language—Communication. An International Handbook on Multimodality and Human Interaction*; De Gruyter Mouton: Berlin, Germany, 2014; Volume 2.
7. Kendon, A. *Gesture—Visible Action as Utterance*; Cambridge University Press: Cambridge, UK, 2004.
8. McNeill, D. *Gesture and Thought*; University of Chicago Press: Chicago, IL, USA, 2005.
9. Kendon, A. Some relationships between body motion and speech: An analysis of an example. In *Studies in Dyadic Communication*; Siegman, A.W., Pope, B., Eds.; Elsevier: New York, NY, USA, 1972; pp. 177–210.
10. Kendon, A. Gesticulation and speech: Two aspects of the process of utterance. In *Nonverbal Communication and Language*; Key, M.R., Ed.; Mouton: The Hague, The Netherlands, 1980; pp. 207–227.
11. Bressem, J.; Ladewig, S.H. Rethinking gesture phases: Articulatory features or gestural movement? *Semiotica* **2011**, *1*, 53–91. [CrossRef]
12. Kendon, A. How gestures can become like words. In *Crosscultural Perspectives in Nonverbal Communication*; Poyatos, F., Ed.; Hogrefe: Toronto, ON, Canada, 1988; pp. 131–141.

13. McNeill, D. *Hand and Mind. What Gestures Reveal about Thought*; University of Chicago Press: Chicago, IL, USA, 1992.
14. Müller, C. Gesture and Sign: Cataclysmic Break or Dynamic Relations? *Front. Psychol.* **2018**, *9*, 1651. [CrossRef] [PubMed]
15. Ekman, P.; Friesen, W. Head and body cues in the judgment of emotion: A reformulation. *Percept. Mot. Skills* **1967**, *24*, 711–724. [CrossRef]
16. Ekman, P.; Friesen, W. The repertoire of nonverbal behaviour: Categories, origins, usage, and coding. *Semiotica* **1969**, *1*, 49–98. [CrossRef]
17. Ekman, P.; Sorenson, E.; Friesen, W. Pan-cultural elements in facial displays of emotion. *Science* **1969**, *164*, 68–88. [CrossRef]
18. Ekman, P. Body position, facial expression and verbal behaviour during interviews. *J. Abnorm. Soc. Psychol.* **1964**, *68*, 295–301. [CrossRef]
19. Ekman, P. Movements with precise meaning. *J. Commun.* **1976**, *26*, 14–26. [CrossRef]
20. Ekman, P. About brows: Emotional and conversational signals. In *Human Ethology: Claims and Limits of a New Discipline*; von Cranach, M.; Foppa, K., Lepenies, W., Ploog, D., Eds.; Cambridge University Press: Cambridge, UK, 1979; pp. 169–202.
21. Kendon, A. Some functions of the face in a kissing round. *Semiotica* **1975**, *15*, 299–334. [CrossRef]
22. McNeill, D. So you think gestures are nonverbal? *Psychol. Rev.* **1985**, *92*, 350–371. [CrossRef]
23. Bavelas, J.B.; Chovil, N. Faces in dialogue. In *The Psychology of Facial Expression*; Russell, J., Fernandez-Dols, J., Eds.; Cambridge University Press: Cambridge, UK, 1997; pp. 334–346.
24. Bavelas, J.B.; Chovil, N. Visible acts of meaning: An Integrated Message Model for language in face-to-face dialogue. *J. Lang. Soc. Psychol.* **2000**, *19*, 163–194. [CrossRef]
25. Bavelas, J.B.; Chovil, N. Hand gestures and facial displays as part of language use in face-to-face dialogue. In *Handbook of Nonverbal Communication*; Sage: Thousand Oaks, CA, SUA, 2006; pp. 97–115.
26. Bavelas, J.B.; Gerwing, J. Conversational hand gestures and facial displays in face-to-face dialogue. In *Social Communication*; Fiedler, K., Ed.; Psychology Press: New York, NY, USA, 2007; pp. 283–308.
27. Bavelas, J.B.; Gerwing, J.; Healing, S. Hand gestures and facial displays in conversational interaction. In *Oxford Handbook of Language and Social Psychology*; Holtgraves, T., Ed.; Oxford University Press: New York, NY, USA, 2014; pp. 111–130.
28. Bavelas, J.B.; Gerwing, J.; Healing, S. Including facial gestures in gesture-speech ensembles. In *From Gesture in Conversation to Visible Action as Utterances*; Seyfeddinipur, M., Gullberg, M., Eds.; John Benjamins: Amsterdam, The Netherlands, 2014; pp. 15–34.
29. Bavelas, J.B.; Chovil, N. Some pragmatic functions of conversational facial gestures. *Gesture* **2018**, *17*, 98–127. [CrossRef]
30. Wagner, P.; Malisz, Z.; Kopp, S. Gesture in Speech and Interaction: An Overview. *Speech Commun.* **2014**, *57*, 209–232. [CrossRef]
31. Rauscher, F.H.; Krauss, R.; Chen, Y. Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychol. Sci.* **1996**, *7*, 226–231. [CrossRef]
32. Hadar, U.; Butterworth, B. Iconic gestures, imagery, and word retrieval in speech. *Semiotica* **1997**, *115*, 147–172. [CrossRef]
33. Krauss, R.M.; Hadar, U. The role of speech-related arm/hand gestures in word retrieval. In *Gesture, Speech, and Sign*; Oxford University Press: Oxford, UK, 1999. [CrossRef]
34. Krauss, R.M.; Chen, Y.; Gottesman, R. Lexical gestures and lexical access: A process model. In *Language and Gesture*; McNeill, D., Ed.; Cambridge University Press: Cambridge, UK, 2000; pp. 261–283.
35. Kita, S. How representational gestures help speaking. In *Language and Gesture*; McNeill, D., Ed.; Cambridge University Press: Cambridge, UK, 2000; pp. 162–185.
36. Alibali, M.; Kita, S.; Young, A. Gesture and the process of speech production: We think, therefore we gesture. *Lang. Cogn. Process.* **2000**, *15*, 593–613. [CrossRef]
37. Hostetter, A.; Alibali, M. Raise your hand if you're spatial: Relations between verbal and spatial skills and gesture production. *Gesture* **2007**, *7*, 73–95. [CrossRef]
38. Kopp, S.; Bergmann, K.; Kahl, S. A spreading-activation model of the semantic coordination of speech and gesture. In Proceedings of the 35th Annual Meeting of the Cognitive Science Society (COGSCI 2013), Berlin, Germany, 31 July–3 August 2013.
39. Bergmann, K.; Kahl, S.; Kopp, S. Modeling the semantic coordination of speech and gesture under cognitive and linguistic constraints. In Proceedings of the International Conference on Intelligent Virtual Agents (IVA 2013), Edinburgh, UK, 29–31 August 2013.
40. Goldin-Meadow, S.; Brentari, D. Gesture, sign and language: The coming of age of sign language and gesture studies. *Behav. Brain Sci.* **2017**, *40*, e46. [CrossRef] [PubMed]
41. Butterworth, B.; Beattie, G. Gesture and Silence as Indicators of Planning in Speech. In *Recent Advances in the Psychology of Language: Formal and Experimental Approaches*; Campbell, R.N., Smith, P.T., Eds.; Springer: Boston, MA, USA, 1978; pp. 347–360. [CrossRef]
42. Schegloff, E. On some gestures' relations to speech. In *Structures of Social Action*; Atkinson, J., Heritage, J., Eds.; Cambridge University Press: Cambridge, UK, 1984; pp. 226–296.
43. Morrels-Samuels, P.; Krauss, R.M. Word familiarity predicts temporal asynchrony of hand gestures and speech. *J. Exp. Psychol. Hum. Learn. Mem.* **1992**, *18*, 615–622. [CrossRef]
44. Shattuck-Hufnagel, S. Toward an (even) more comprehensive model of speech production planning. *Lang. Cogn. Neurosci.* **2019**, *34*, 1202–1213 [CrossRef]
45. Caschera, M.C.; Grifoni, P.; Ferri, F. Emotion classification from speech and text in videos using a multimodal approach. *Multimodal Technol. Interact.* **2022**, *6*, 28. [CrossRef]

46. Eibl-Eibesfeldt, I. Ritual and Ritualization from a Biological Perspective. In *Non-Verbal Communication*; Hinde, R., Ed.; Cambridge University Press: Cambridge, UK, 1972; pp. 297–312.
47. Noordewier, M.K.; van Dijk, E. Surprise: Unfolding of facial expressions. *Cogn. Emot.* **2019**, *33*, 915–930. [CrossRef] [PubMed]
48. Schützwohl, A.; Reisenzein, R. Facial expressions in response to a highly surprising event exceeding the field of vision: A test of Darwin's theory of surprise. *Evol. Hum. Behaciour* **2012**, *33*, 657–664. [CrossRef]
49. Darwin, C. *The Expression of Emotion in Man and Animals*; John Murray: London, UK, 1872.
50. Wierzbicka, A. The semantics of human facial expression. *Pragmat. Cogn.* **2000**, *8*, 147–183. [CrossRef]
51. Gast, V. Eyebrow raises as facial gestures: A Study based on American late night show interviews. In *Language Use and Linguistic Structure. Proceedings of the Olomouc Linguistics Colloquium 2021*; Markéta Janebová, J.E., Veselovská, L., Eds.; Palacký University Olomouc: Olomouc, Czech, 2022; pp. 259–279.
52. Chovil, N. Social determinants of facial displays. *J. Nonverbal Behav.* **1991**, *15*, 141–154. [CrossRef]
53. Chovil, N. Discourse-oriented facial displays in conversation. *Res. Lang. Soc. Interact.* **1991**, *25*, 163–194. [CrossRef]
54. Kim, J.; Cvejic, E.; Davis, C. Tracking eyebrows and head gestures associated with spoken prosody. *Speech Commun.* **2014**, *57*, 317–330. [CrossRef]
55. Flecha-García, M.L. Eyebrow Raising in Dialogue: Discourse Structure, Utterance Function, and Pitch Accents. Ph.D. Thesis, University of Edinburgh, Edinburgh, UK, 2006.
56. Flecha-García, M.L. Eyebrow raises in dialogue and their relation to to discourse structure, utterance function and pitch accents in English. *Speech Commun.* **2010**, *52*, 542–554. [CrossRef]
57. Brown, G.; Anderson, A.; Yule, G.; Shillock, R. *Teaching Talk*; Cambridge University Press: Cambridge, UK, 1983.
58. Anderson, A.; Bader, M.; Bard, E.; Boyle, E.; Doherty, G.; Garrod, S.; Isard, S.; Kowtko, J.; McAllister, J.; Miller, J.; et al. The Hcrc Map Task Corpus. *Lang. Speech* **1991**, *34*, 351–366. [CrossRef]
59. Cavé, Christian, I.G.; Santi, S. Eyebrow movements and voice cariations in dialogie situations: An experimental investigation. In Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP2002), Denver, CO, USA, 16–20 September 2002.
60. Guaïtella, I.; Lagrue, B.; Cavé, C. Are eyebrow movements linked to voice variations and turn-taking in dialogue? *Lang. Speech* **2009**, *52*, 207–222. [CrossRef]
61. Danner, S.G.; Krivokapić, J.; Byrd, D. Co-speech movement in conversational turn-taking. *Front. Commun.* **2021**, *6*, 779814. [CrossRef]
62. Danner, G.S. Effects of Speech Context on Characteristics of Manual Gesture. Ph.D. Thesis, University of Southern California, Los Angeles, CA, USA, 2017.
63. Bolinger, D. Intonation and gesture. In *American Speech*; Duke University Press: Durham, NC, USA, 1983; Volume 58, pp. 156–174.
64. Cavé, C.; Guaïtella, I.; Bertrand, R.; Santi, S.; Harlay, F.; Espesser, R. About the relationship between eyebrow movements and F0 variations. In *Proceedings of the 4th International Conference on Spoken Language Processing*; Isardi, H.T.B.W., Ed.; University of Delaware and Alfred I. duPont Institute: Newark, DE, USA, 1996; pp. 2175–2178.
65. Schade, L.; Schrumpf, M. The coordination of eyebrow movement and prosody in affective utterances. In *Ein Transdisziplinäres Panoptikum. Aktuelle Forschungsbeiträge aus Dem Wissenschaftlichen Nachwuchs der Universität Bielefeld*; Pawlak, O.M., Rahn, F.J., Eds.; Springer: Heidelberg, Germany, 2021; pp. 96–106.
66. Amos, B.; Ludwiczuk, B.; Satyanarayanan, M. *OpenFace: A General-Purpose Face Recognition Library with Mobile Applications*; Technical Report, CMU-CS-16-118; CMU School of Computer Science: Pittsburgh, PA, USA, 2016.
67. Boersma, P.; Weenink, D. Praat: Doing Phonetics by Computer [Computer Program]. Version 6.0.37. 2018. Available online: http://www.praat.org/ (accessed on 14 March 2018).
68. Berger, S.; Zellers, M. Multimodal prominence marking in semi-spontaneous YouTube monologs: The interaction of intonation and eyebrow movements. *Front. Commun.* **2022**, *7*, 132. [CrossRef]
69. Wells, J. *English Intonation: An Introduction*; Cambridge University Press: Cambridge, UK, 2006.
70. Pierrehumbert, J. The Phonology and Phonetics of English Intonation. Ph.D. Thesis, UILC, Bloomington, IN, USA, 1980.
71. Bulat, A.; Tzimiropoulos, G. How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks). In Proceedings of the International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
72. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022.
73. Kuznetsova, A.; Brockhoff, P.B.; Christensen, R.H.B. lmerTest Package: Tests in Linear Mixed Effects Models. *J. Stat. Softw.* **2017**, *82*, 1–26. [CrossRef]