The Charles - A new sensor device for measuring body language and stress in Speech Communication

Vidar Freyr Gudmundsson[1], *Ïo Valls-Ratés*[2], *Oliver Niebuhr*[1]

[1]*University of Southern Denmark,* [2]*Universitat Pompeu Fabra*
vigud20@student.sdu.dk

Abstract: We present The Charles, a new 10-channel sensor belt system for measuring and recording body-language and temperature-related stress signals that a speaker produces, for example, in a public-speaking context. Key results of an initial proof-of-concept test are summarized, carried out with 8 male and female speakers who gave a speech in L2 English. We found that the speakers' arm, hand, and torso movements show many strong correlations with prosodic patterns, albeit with clear differences in terms of laterality and speaker gender. The findings are discussed with respect to the applicability of The Charles and its measures and directions of future research.

1. Introduction

1.1 Body Language: An Emergent Research Field of the Speech Sciences

Decades ago, research in speech sciences was primarily concerned with the forms and functions of sound segments (phones). With the development of intonational phonologies, there was a rapid expansion of the research perspective into the area of suprasegmentals or prosody [1]. Prosody research initially focused primarily on fundamental frequency (f0) and duration patterns. Only later did the perspective become four-dimensional [2] or, actually, five-dimensional, if the interplay of phones and prosodies [3] is considered a research dimension in its own right.

The speech sciences are currently again on the threshold of the next significant expansion of their research perspective. This further expansion concerns the “second channel” of non-verbal communication, i.e. the signals of mimic and gestures or, in short, the “body language” that is conveyed in parallel to the other non-verbal channel of speech prosody.

Brown & Prieto [4] recently emphasized that sociopragmatics, a field still dominated by research on words, cannot do without taking prosody and body language into account. Brown & Prieto show, supported by many examples, “that prosody is closely integrated with gesture both at the temporal level and in the kinds of pragmatic meanings that these two systems are used to encode” (p.430), i.e. attitude, stance, (im)politeness, irony, etc. In the Introduction to their special issue, Wagner et al. [5] summarize the multi-faceted interplay of prosody and body language across the world’s languages. Beyond the summary itself, the above-mentioned threshold at which speech sciences currently stand with regard to body language becomes visible in [5] in the list of references. It includes twice as many papers from the 21st as from the 20th century. Wagner et al. conclude in this context: “we expect that our future understanding will profit from the, now widespread, availability of technical tools such as annotation software and affordable solutions for building multimodal corpora” (p. 227). It is precisely this expectation that we want to contribute to with our developed sensor-belt system “The Charles”.

1.2 “The Charles”: Background and Features

Named after Charles Darwin, who was among the first to write about body language [6], “The Charles” is a motion capture device (mocap) built for body-language research. Many commercially available mocap systems exist, but they are mainly built for animating computer-generated imagery (CGI) and have, if at all, only a secondary focus on research. Moreover, commercially available mocap systems are usually very expensive and require highly trained personnel to operate it and, in many cases, specialized laboratory environments. The Charles project seeks to make these processes more cost-effective and mobile. Tiny and inexpensive micro-electro-mechanical system (MEMS) sensor technology is used that makes research into body language or related fields such as ergonomics and biomechanics more widely accessible and affordable for universities and other research institutions.

Furthermore, The Charles uses Inertial Measurements Units (IMU), whereas many other existing systems are based on optical solutions. The latter require a number of cameras and lights, which severely limit the systems' mobility and entail long setup times; and for linguistic fieldwork, an important activity in the context of globalization [7], such systems are completely unsuitable. In addition, a sampling rate of 24 frames per second (fps), i.e. one measurement per 41.7 milliseconds (ms), is commonly used in cinematic feature movies. While this is enough to make the illusion of seamless motion to the human eye, higher framerates are needed to study body language and its coordination with speech prosody. Speakers are able to coordinate their articulation and phonation with striking precision. Research on pitch-accent alignment shows this very clearly [8-9]. Gestures are also timed to within a few milliseconds with phonatory landmarks [5, 10-11]. Thus, for investigating human body language, in particular its embedding in the overall system of phonetic movement patterns, a higher temporal resolution is required than is usual in video playback.

Similar the temporal-resolution issue, sensitivity to interference is a concern in research, but less so in computer animation, where a lot of post processing occurs. There is a high risk of marker occlusion in optical mocap systems, especially when the light conditions are not ideal (which they rarely are "in the field"). Many mocap systems contain algorithms to extrapolate the missing frames and trajectories of body parts, but this is arguably not ideal in science.

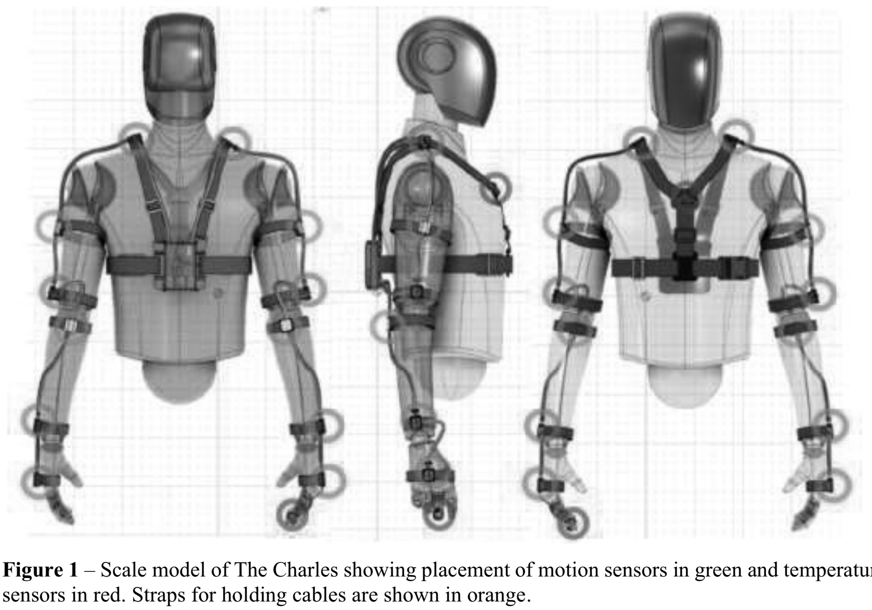

Figure 1 – Scale model of The Charles showing placement of motion sensors in green and temperature sensors in red. Straps for holding cables are shown in orange.

A set of wired IMU units like those used in The Charles is basically capable of recording 200 fps or more, depending on the speed of the microcontroller unit (MCU) and the number of sensors used. Thus, in principle, The Charles could provide a time resolution of 5 ms or less; enough for an in-depth analysis of body language and its interplay with all other verbal and non-verbal communication signals. Its 8 sensors are placed along the body with special emphasis on measuring communicative gestures, see Figure 1. For example, there are separate sensor units at the wrist *and* at the (back of the) hand, taking into account the major and partly separate contributions that these two parts of the arm make to communicative gestures. Moreover, each of the system's 8 encapsulated sensors combines the 3D measurements from an accelerometer and a gyroscope, which allows a complex analysis of gestures from different points of view. Another special feature is that The Charles also records a synchronization signal, which later allows a time-aligned analysis of all measurements with the acoustic speech signal.

By all measurements we mean more than just movements. In addition to the 8 motion sensors, The Charles features two high precision temperature sensors: one at the fingertip and one at the wearer's chest, see Figure 1. Research has shown that the temperature difference between distal and proximal locations increases in stressful situations [12]. Due to a change in blood flow, originally meant to prepare the body for life-threatening circumstances, the extremities get colder while the trunk gets warmer. The temperature at the fingertips can fall by $\Delta T=-2.7°C$ as compared to the torso temperature. The Charles is thus able to determine the wearer's stress level, for example, in public-speaking situations, and it allows researchers to combine this data with the analysis of body language. To the best of our knowledge, our sensor-belt system is currently the only one that can measure stress in the form of temperature differences. A future version of The Charles will also include a heart-rate sensor on the chest belt as well as a humidity sensor at the wearer's back to measure perspiration due to physical or mental stress.

Regarding comfort and the related issue of ecological validity (i.e. the measurement bias caused by the intrusiveness of the device), The Charles weighs less than 300 g and can be worn under normal clothing. The cables that connect the sensors and the central processing unit are securely held in place by Velcro straps. To make cables as unobtrusive as possible, special cable holding harnesses were designed for each forearm and upper arm. The main enclosure that hosts a single 9V battery and the main printed circuit board is subtly placed at the user's lower back, fixed by a wearable harness of elastic straps that can be adjusted to fit any body type. The central processing unit is an Arduino MCU. It switches between the sensors and collects measurements while sending the data over to a Bluetooth module. The data are then picked up by an app on an Android smartphone in the form of a timestamped comma-separated-values file (CSV). The use of mobile phone application was to enhance the mobility of the system, allowing recordings of data to be made in remote locations without the use of computers or Wi-Fi [7]. The data sent via Bluetooth are gyroscope measurements, accelerometer measurements and temperature measurements along with the relative timing of each reading (the latter can in a following stage be time-aligned with the acoustic speech signal).

1.3 Motivation and Aim

The main reason for using a device such as The Charles for body-language research is to understand how humans move when they talk, i.e. how emotions, sociopragmatic meanings, and other linguistic functions (topic, focus, etc.) manifest themselves in movements, also in relation to those of articulation and phonation. Beyond informing models and theories [5], the obtained knowledge has implications for several practical applications. In broad terms these are:

- ➢ Teaching humans to perform better in public speaking or to learn the body-language inventory of a foreign language.
- ➢ Teaching robots or virtual humans to act more human-like. This includes but is not limited to robotic caretakers, computer-generated avatars in virtual reality or gaming, and robots that do greeting service.
- ➢ Teaching AI to better understand humans. This includes but is not limited to supporting speech recognition by taking into account non-verbal gestures, estimating people's moods and feelings for health-related purposes, and aiding security surveillance systems by identifying erratic behavior cues.

Building on the works of [13-15], the present paper focuses on public speaking. We present the results of a pilot data collection that had two objectives: first, to conduct a proof-of-concept test with The Charles; and second, to collect initial movement data in a public-speaking scenario as a basis for data-driven hypotheses in subsequent studies. Proof of concept is provided when speakers can comfortably present with The Charles, while the system continuously collects movement data, whose consistency and plausibility – with a view to the second objective – manifests itself, for example, in revealing correlations between movements and prosodies.

2. Method

2.1. Participants

A total of 8 participants (50% male and 50% female) took part in the study. They were all experienced speakers and recruited at University of Southern Denmark (SDU). Table 1 summarizes key parameters of our speaker sample. Note that, like in the population as whole [16], the vast majority of our sample speakers were right-handed.

Table 1 – Summary of the speakers and the recorded speeches they gave

SPEAKER	SEX	AGE	PRES TIME	IPU COUNT
KAN	f	23	03:58	87
MSV	f	45	05:25	127
IVA	f	34	07:46	285
KFI	f	54	09:20	189
ONI	m	44	10:30	230
KPA	m	59	06:19	147
MOC	m	53	03:29	86
AKO	m	42	06:36	158

2.2. Measures

In preparation for the recordings, the participants were fitted with the Charles sensor belt system. In addition, they were fitted with the Muse II EEG headset [17]. A lavalier microphone was fixed to the sensor belt system, just below the neck but invisible to the speaker. The acoustic speech signal was recorded with a Zoom H4N at 48 kHz, 24 bit. In addition, all speakers stood on the Sensfloor system [18] when presenting. This pressure-sensitive mat with 64 sensors per m² recorded the contacts of the speaker's feet with the floor.

Thus equipped (see Fig.2), the following data was collected from the speakers: The 3D movement characteristics of arm, hand, and shoulder in the form of the gyroscope and accelerometer data at all 8 sensor positions of The Charles (x, y, z axes); the 3D head-movement characteristics of the speaker (x, y, z axes) based on the gyroscope values of the Muse II headset; the EEG activity in the speaker's prefrontal and temporal cortex areas, represented by two key positions each (AF7, AF8, TP9, TP10); the area walked on and the weight distribution of the feet while standing and, finally, the acoustic speech signal (time-aligned to all other signal sources). On top of that, participants were video-recorded throughout their speeches to be able to further analyze their presentation performances in future perception studies.

Only the gyroscope data of the Charles sensor belt system along the three dimensions of the x, y, and z axis were analyzed for this pilot study. The overall picture of the findings is reported in a later paper.

2.3. Procedure

The study was conducted at the University of Southern Denmark (SDU) in Sonderborg. All participants carried out the same task, i.e. giving a brief speech inside the Acoustics Lab in front of a video camera and a small audience, see Figure 2. The recordings took place in connection with an international conference held at SDU (TAI 2021). Thus, all participants were familiar with their speech and were able to perform it fluently, because they had practiced it for the conference or had just held it at the conference the same day or the day before. Each speaker performed the speech individually and in L2 English (a language used by all speakers in a daily basis). Individual presentation times ranged from 03:29 to 10:30 minutes, see Table 1.

Figure 2 – Snapshots of MSV's, KPA's, and MOC's presentations, extracted from the video recordings.

2.4. Data analysis

The data analysis was guided by the acoustic signal. That is, semi-automatic annotations were conducted that divided the recorded presentations into prosodic phrases, more precisely into inter-pausal units (IPUs) defined as sections of speech with audible silent pauses (> 200 ms) at both ends. Table 1 summarizes how many of these IPUs have been annotated per speaker. Basic prosodic measurements were taken for each IPU. The measurements were carried out automatically using ProsodyPro [19]. Implausible values were manually remeasured or removed. The measurements comprised three local fundamental-frequency (f0) points, i.e. minimum f0, maximum f0, and final f0. In addition, the f0 range (excursion size) was determined as well as the average f0 level and the average intensity level (RMS). With regard to the prosodic dimension of timing, the duration of the IPU was measured, and furthermore the maximum f0 velocity within the IPU and the location of the f0 maximum within the IPU (relative to the IPU duration).

Regarding the Charles data associated with each IPU, we were, for the purposes of the present pilot study, less interested in spatial locations and exact directions of movement but more in how prosodic levels and dynamics can be mapped onto those of the body. Therefore, we focussed on the gyroscope values measured at each of the 8 sensor positions. All three axes were taken into account. From the speaker's point of view, the x-axis runs along the front-back dimension, i.e. along the speaker's sightline. The y-axis runs sideways along the speaker's left-right dimension, also horizontally. The z-axis runs along the top-bottom dimension, i.e. between the floor and the ceiling from the speaker's point of view. The gyroscopes measured rotations around these virtual axes, see also Figure 1 in [5], p. 212.

In the typical baseline position that the arms of speakers assume in presentations (see also Fig. 2), the gyro measurements along the three axes can be interpreted – roughly simplified and generalized – as follows. For hands and arms, rotations along the x-axis correspond to raising and lowering movements, as in prototypical up-down beat gestures. The rotation along the y-axis can be attributed to rotational movements of the hands, for example, with regard to turning the palm(s) upward or downward or inward or outward. Rotations along the z-axis correspond to lateral movements, such as opening and closing gestures of the arms in front of the body. As for the speaker's shoulders, gyroscope deflections along the x-axis indicate a lateral swaying of the body – or a shrug of the shoulders; y-axis deflections correspond to a front-to-back see-saw motion; and z-axis deflections result mainly from a left-right rotation of the torso, which is perhaps the clearest reflection of an audience-oriented presentation behavior.

From the movement data obtained along each axis, two measures were formed per IPU: the mean value and the skewness. The mean provides a simple overview of the speed with which the speaker has moved/rotated his or her hands, arms, and shoulders within the respective IPU. Moreover, in view of the "speed-amplitude relation" that has been found in many studies across species and which seems to be a general characteristic of biological movement patterns [20], speed translates into movement amplitude. In other words, we can assume that higher rotation speeds are to a large degree indicative of larger movements.

The choice of skewness as the second measure also refers to the "speed-amplitude relation". Skewness represents the degree to which the measurements at each sensor are distorted to either smaller (negative) or higher (positive) values. A distortion towards smaller values means positive skewness, a distortion towards higher values negative skewness. Speakers are biological organisms and therefore unable to keep their arms and hands completely still while speaking. There will always be a noise floor in the form of minor movements with correspondingly minor rotation speeds. For confident speakers who, in a controlled manner, perform meaningful gestures that stand out clearly from this background noise of involuntary minor movements, clearly positive skewness values will emerge. That is, the few fast and large rotations related to these meaningful gestures are embedded into the vast majority of slow and small rotations that constitute of noise floor. By contrast, the more speakers are nervous and fidgety when speaking, i.e. the more uncontrolled and meaningless the gestures are that they perform, the smaller are the skewness values they create. In extreme cases, skewness can become negative. In summary, we use skewness here as a measure of confidence and the goodness of the contrast that the speaker is able to create between the inevitable movement noise on the one hand and a few, controlled and meaningful gestures on the other. The higher the skewness value the better does the speaker perform in that respect.

3. Results and Discussion

Gestural characteristics in presentation behavior and their inter-individual differences cannot be dealt with in detail in this short paper. As described earlier, the aim of this study is to determine whether and how movements along the x-, y- and z-axes are related to prosodic features – as a proof-of-concept test of The Charles and as a basis for hypotheses and generalizations in subsequent studies. To that end, the results section focuses on key correlations between the measured prosodic features on the one hand the mean amplitude and skewness of body movements measured at the 8 sensor positions on the other. Pearson product-moment correlations were conducted, with p-values being corrected for multiple testing based on the Benjamini-Hochberg procedure. The sample size in all correlations was $N = 1{,}309$ ($df = 1{,}307$).

Five selected key correlation patterns are presented below, all based on 'strong' correlations in the sense of Cohen's effect sizes [21], i.e. correlations with p-values < 0.01 and Pearson's r values > 0.5. Note that, because the results are complex and difficult to grasp in isolation, we combined the presentation of the results with their discussion.

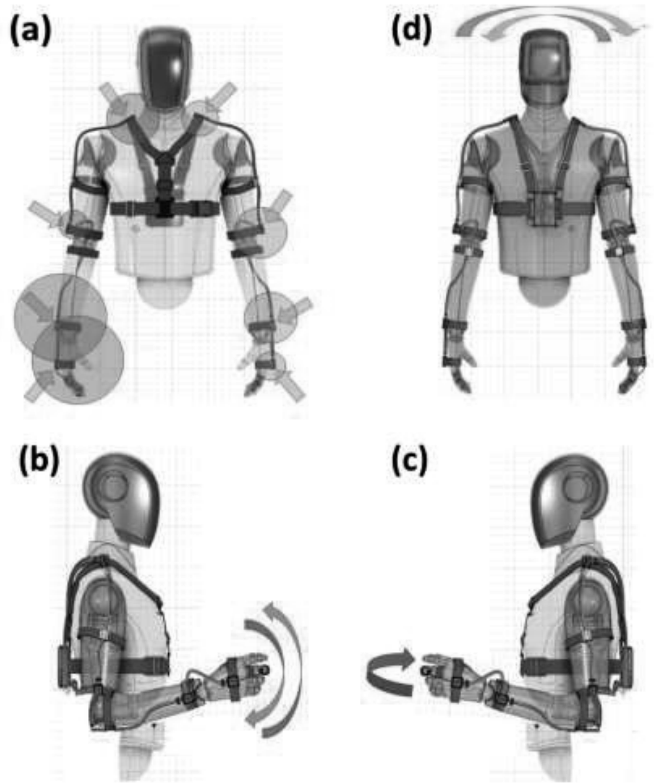

Figure 3 – Illustration key results; (a): circle sizes indicate how many prosodic features are correlated with movements at the respective sensor location; (b)-(c): positive correlations of movements with f0 excursion size; (d): left-right torso rotation, correlated positively with f0 excursion size and intensity level and negatively with f0 fall and IPU duration. Blue, orange, and green symbols mark the speaker's right side, left side, or the whole torso, respectively.

The first and most obvious result was a clear lateral asymmetry in the significant correlations between body movements and prosodic characteristics, see Figure 3(a): Of all correlations between prosody and movement means, 87 % (20 out of 23) related to the speaker's right side and only 13 % to the left side. A similar picture emerged for the skewness measure; 75 % of the correlations with prosodic features related to the right side and only 25 % to the left side. The overall amount of correlations emphasizes, in line with many previous studies [4-5, 10-11], how remarkably closely body movements are coordinated with prosodic features. The lateral asymmetry within these correlations further suggests that speakers primarily used their right hand, arm and shoulder for this coordination, i.e. for the execution of controlled, meaningful gestures. This interpretation is underpinned by the fact that the vast majority of our recorded speakers was in fact right-handed.

Nevertheless, speakers of course moved both sides of their body while speaking. In this connection, a second key result emerged: The f0 excursion size was consistently positively correlated with movement means, but along different axes on the two sides of the body. On the right-hand side, the f0 excursion size was positively correlated with hand/forearm movements along the z-axis; whereas on the left-hand side, it was positively correlated with hand/forearm movements along the x-axis. So, while faster and correspondingly more extensive movements generally coincided with a more variable speech melody (in accord with previous studies

[5,21]), the data seem to indicate that it were mainly vertically executed up/down beat gestures of the left hand and forearm (Fig. 3b) as well as laterally executed opening/closing gestures of the right hand and forearm (Fig. 3c) that were associated with this more variable melody.

Third, not all prosodic features were correlated in the same way with the means and skewness values of the speaker's movements. Movement means were mostly (to 69 %) linked to f0 characteristics or the intensity level. The skewness values were mostly (to 66 %) linked to IPU duration or features of f0-peak timing. In other words, how fast and how much the speaker moved the body was primarily associated with melody and timbre; whereas how selectively and controlled gestures were carried out was primarily associated with timing or, more generally, the syntagmatic structuring of prosody.

Fourth, the shoulder-sensor measurements also correlated with prosodic features. More extensive z-axis movements were associated with greater f0 excursion sizes as well as with deeper phrase-final f0 falls, a higher speech intensity level, and more frequent silent pauses in terms of shorter IPU durations (Fig. 3d). This overall prosodic setting is characteristic of charismatic speakers like Steve Jobs [24], which fits well with the above description of z-axes shoulder rotations as the perhaps clearest reflection of audience-oriented presentation behavior. By contrast, shoulder movements along the x- and y-axes were correlated with prosodic settings too, but with less beneficial ones for public speaking. This was especially true for more extensive movements along the y-axis, i.e. the back-to-front see-saw motion of the body, captured by the shoulder sensors. The stronger such a see-saw motion, the smaller the f0 deflection magnitude became and the earlier (i.e. higher) the phrase-final f0 falls ended. The mean f0 level also increased. This means that the presenter's speech melody was monotonized at a high pitch level. Larger movements along the x-axis, indicative of laterally swaying body movements or shoulder shrugs, showed no connection to intonation, but were highly correlated with a lower intensity level, i.e. with a softer voice, and with fewer silent pauses and hence longer IPU durations.

Closer examinations of the previous, fourth key result finally revealed the fifth one: the movement patterns recorded at the shoulder sensors were to some extent gender-specific. Compared to the male speakers, the female speakers showed significantly more and stronger y-axis movements while presenting, i.e. back-to-front see-saw motions of the body, whereas the male speakers showed more and stronger z-axis movements, i.e. leftward and rightward rotations of the body (Fig. 3d). Furthermore, in terms of skewness, we found across almost all sensors higher and/or more positive skewness values for the male than for the female speakers. Thus, if we interpret the skewness data as an indirect reflection of speaker confidence defined by the ability to perform few, controlled, and meaningful gestures against a background of inevitable low-amplitude motion noise, then our data indicate higher levels of confidence and control among the male speakers in comparison to the female speakers.

4. Conclusion and Outlook

The measurement of body language under different situational and communicative contexts has the potential to become a new major research field in the speech sciences, after prosody or in close connection with it. This rapidly developing research field calls for new tools and technologies for collecting (and annotating) data, cf. [5]. Against this background, our paper introduces a new tool that can make a major contribution to body language exploration, both through its mobility and through its accessibility in terms of an affordable, easy-to-use system. The tool is the sensor belt system called The Charles. It offers a 10-channel recording of body movements and temperature-related stress signals, which are transmitted as a CSV file via Bluetooth to the user's smartphone, time-aligned to the acoustic speech signal. The Charles can be conveniently controlled via a specifically developed app on the user's smartphone.

We collected pilot data for an initial proof-of-concept test of the system and for empirically based hypothesis building in future studies. Based on our results, The Charles passed the proof-of-concept test. We found a number of significant correlations of the recorded sensory gyroscope data with acoustic-prosodic parameter changes. If the gyroscope data recording had not been continuous or not time-aligned with the acoustic speech signal – or if the measurements had been erratic and not recurrently associated with prosodic patterns, then no significant correlations would have been found; in particular not in the reported quantity and strength, i.e. with $r > 0.5$ (and partly $r > 0.8$) for a sample size of $N = 1.309$. Insofar, we can assume that The Charles worked flawlessly and delivered a coherent and plausible stream of measurements. This conclusion is supported by the fact that the links that emerged between a more extensive body language and a more variable prosody or intonation were also found in previous studies [5].

In addition, we have gained new insights that are worth being further explored in the future. This includes, for example, that the arm, hand, and torso movements show both lateral and gender-specific differences in the way they are linked to prosody. The new measure of skewness of movement/rotation amplitudes used here should also be examined more closely in follow-up studies for its interpretability with regard to the speaker's confidence and control.

Future research on the link between body movements and prosody features should additionally address the perception of the speaker, i.e. to what degree there is a positive correlation between a more variable or extensive body language and more charismatic speaker attributes or a stronger persuasiveness of the message. For example, Maricchiolo et al. [22] concluded that speakers who use more body language are perceived as more composed, competent, and likeable than those who use less body language. The Charles could break down such effects in more detail. Also, we would like to use The Charles for a closer examination of how levels of public-speaking anxiety affect non-verbal communication signals, for example, with regard to their variability, dynamics, and amplitude. We know from studies that the anxiety experienced by speakers is reflected in their f0, duration or intensity patterns [23]. The exact body-language exponents of different levels of experienced anxiety and their relation to the perceived anxiety of the speaker on the part of the audience, however, are still far from being well understood. In summary, we believe that our pilot data and the new device presented here are a promising starting point for fleshing out the relationships between prosody and body language in various ways as well as over and above delivering public speeches.

Acknowledgements

This work was made possible by the Fabrikant Mads Clausen Foundation whose generous funding allowed us to design, build and test the new sensor belt system "The Charles". Moreover, we are greatly indebted to Igor Pawel Smorag for his many hands-on contributions to the project, including the programming of the system's smartphone app.

References

[1] LADD, D. R.: *Intonational phonology.* Cambridge University Press, 2008.

[2] CAMPBELL, N., & P. MOKHTARI: Voice quality: the 4th prosodic dimension. *15th International Congress of Phonetic Scicnes, Barcelona, Spain,* pp. 2417-2420, 2003.

[3] MÜCKE, D., M. GRICE, & T. CHO: More than a magic moment–Paving the way for dynamics of articulation and prosodic structure. *Journal of Phonetics, 44*, pp. 1-7, 2014.

[4] BROWN, L., & P. PRIETO: Gesture and Prosody in Multimodal Communication. In M. HAUGH, D. Z. KÁDÁR, & M. TERKOURAFI (eds.), *The Cambridge Handbook of Sociopragmatics* (pp. 430-453). Cambridge: Cambridge University Press, 2021.

[5] WAGNER, P., Z. MALISZ, & S. KOPP: Gesture and speech in interaction: An overview. *Speech Communication, 57*, pp. 209-232, 2014.

[6] DARWIN, C. R.: The expression of the emotions in man and animals. London: John Murray. 1st edition, 1872.

[7] GIPPERT, J., N. HIMMELMANN, & U. MOSEL (eds.): *Essentials of language documentation.* Berlin: Mouton de Gruyter, 2006.

[8] ATTERER, M., & D.R. LADD: On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. *Journal of Phonetics, 32*, pp. 177-197, 2004.

[9] ARVANITI, A.: Prosodic representations Part I: Segment-to-Tone Association. In A. COHN, C. FOUGERON, & M.K. HUFFMAN (eds.), *The Oxford handbook of Laboratory Phonology*. Oxford: Oxford University Press, 2012.

[10] LEONARD, T., F. & CUMMINS: The temporal relation between beat gestures and speech. *Language and Cognitive Processes, 26*, pp. 1457-1471, 2011.

[11] LOEHR, D. P.: Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory phonology, 3*, pp. 71-89, 2012.

[12] VINKERS, C. H., R. PENNING, J. HELLHAMMER, J.C. VERSTER, J.H. KLAESSENS, B. OLIVIER, & C.J. KALKMAN: The effect of stress on core and peripheral body temperature in humans. *Stress, 16*, pp. 520-530, 2013.

[13] NIEBUHR, O. & A. GUTNYK: Pronunciation engineering: Investigating the link between jaw-movement patterns and perceived speaker charisma using the MARRYS cap. IEEE Xplore (Proc. 3rd International Conference on Electrical, Communication, and Computer Engineering), 1-5, 2021.

[14] TSCHINSE, L.M., A. ASADI, A. GUTNYK, & O. NIEBUHR: Keep on smiling...? On the sex-specific connections between smiling duration and perceived speaker attributes in business pitches. *Proc. 11th International Conference of Speech Prosody, Lisbon, Portugal*, pp. 1-5, 2022

[15] VALLS-RATÉS, I., O. NIEBUHR, & P. PRIETO: Unguided VR public-speaking training enhances your confidence - but does not improve your intonation. *Proc. 11th International Conference of Speech Prosody, Lisbon, Portugal,* pp. 1-5, 2022.

[16] PRICE, M.: The left brain knows what the right hand is doing. *Monitor on Psychology, 40*. http://www.apa.org/monitor/2009/01/brain, 2009.

[17] ARSALAN, A., M. MAJID, A.R. BUTT, & S.M. ANWAR: Classification of perceived mental stress using a commercially available EEG headband. *IEEE Journal of Biomedical and Health Informatics, 23*, pp. 2257-2264, 2019.

[18] LAUTERBACH, C., A. STEINHAGE, & A. TECHMER: Large-area wireless sensor system based on smart textiles. In *International Multi-Conference on Systems, Signals & Devices*, pp. 1-2. IEEE, 2012.

[19] XU, Y. : ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. *Proc. Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), Aix-en-Provence, France*, pp. 7-10, 2013.

[20] NEWELL, K. M., P.A. HANCOCK, & R.N. ROBERTSON: A note on the speed-amplitude function in movement control. *Journal of Motor Behavior, 16*, pp. 460-468, 1984.

[21] COHEN J.: *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic, 1988.

[22] MARICCHIOLO, F., A. GNISCI, M. BONAIUTO, & G. FICCA: Effects of different types of hand gestures in persuasive speech on receivers' evaluations. *Language and cognitive processes, 24*, pp. 239-266, 2009.

[23] WEEKS, J. W., C.Y. LEE, A.R. REILLY, A.N. HOWELL, C. FRANCE, J.M. KOWALSKY, & A. BUSH: "The sound of fear": Assessing vocal fundamental frequency as a physiological indicator of social anxiety disorder. *Journal of Anxiety Disorders, 26*, 811–822, 2012.

[24] NIEBUHR, O., A. BREM, J. MICHALSKY, & J. NEITSCH: What makes business speakers sound charismatic? A contrastive acoustic-melodic analysis of Steve Jobs and Mark Zuckerberg. *Cadernos de Linguistica e Teoria da Literatura*, *1*, pp. 1-40, 2020.