



Integrated Communication Systems and Kinesic Code-Switching in Interpersonal Interaction

Mikael Jensen, Linnéa Moreira Emanuelsson

Department of Applied Information Technology,
University of Gothenburg, SE-412 96 Göteborg, Sweden

Corresponding author: Mikael Jensen, Department of Applied Information Technology, University of
Gothenburg, SE-412 96 Göteborg, Sweden

Abstract: The purpose of the study is to find support for the integrated communication systems hypothesis. To find support for this hypothesis a specific group of participants were selected: bicultural and bilingual individuals that were tested in four different conditions. A new and highly reliable motion capture system together with special software was used to measure gesture velocity. Two working hypothesis were formulated. The secondary hypothesis was verified: the bicultural and bilingual participants change their gestural pattern depending on what language they speak. We call this a kinesic code-switching. The tendency is somewhat stronger in the face-to-face condition compared to the audio only condition. The primary hypothesis was supported in competition with alternative hypotheses: the only hypothesis that can fit all the results in this study, all four test conditions, is the integrated systems hypothesis. The participants most likely use two intertwined communication systems when they communicate in an interpersonal situation.

Keywords: Gesture, Speech, A unified system, Bicultural, Bilingual, Motion capture.

Introduction

Susan Goldin-Meadow (2003, p. 16) state that several “types of evidence lend support to the view that gesture and speech form a single, unified system”. This is an interesting hypothesis. Before we can look closer at the evidence we need to establish a common understanding of what gesture and speech respectively is, since it will affect the possible alternatives (Kendon, 2000).

In a layman’s term gesture is the use of hands and arms (cf. Poggi & Pelachaud, 2008). What about head movements (Kendon, 2004), are they also gestures? Some also say that facial expressions are gestures, facial gestures (Quek et al., 2002), or even more specific like speech gestures. Callan et al. (2003) define speech gestures as “the biological motion of various articulators (e.g. jaw, lips, tongue, larynx) that specify vocal tract shape”. Can we have leg gestures or foot gestures? When it comes to speech it is not fully clear if speech involves only symbolic oral sounds or also other oral sounds like “em” or “uh-huh”. Is speech, whether it is public speaking or in an interpersonal conversation, just talking or the multimodal aspect of it? Does it include turn management (Allwood, 2008) that is using both verbal and nonverbal means to regulate a conversation? If speech is used in a wider sense it would be trivial to say that speech and gesture are unified, they are obviously highly coordinated. In this paper and in the present study we define speech as the oral symbolic and iconic (e.g. onomatopoeic sounds) use of signs and consequently exclude oral indexical signs (e.g. breathing, coughing, and laughing). Gestures are defined as arm, hand, finger and head movements that can be indexical, iconic and symbolic (cf. Kendon, 1981; 2004). Speech gestures are excluded for the simple reason that they can’t exist if we don’t speak and we can’t speak without producing sounds with the speech organ movements. Speech and speech gestures are thus 100 percent unified. It seems acceptable that head movements like nods and shakes can have a symbolic meaning (Matsumoto & Hwang, 2013a). Otherwise, movements that can’t reach a symbolic (or even iconic) level are excluded in this case (posture, leg movements, foot movements, gaze and facial expressions).

There are several types of gestures and some types are less relevant for speech. Ekman and Friesen (1967/1981; also Afifi, 2010) differentiate between emblems, illustrators, affective display, regulators and adaptors. *Emblems* are gestures that have a symbolic meaning and can easily be translated into a word. They can be used independent of speech as well as together with speech. *Illustrators* are speech dependent, that is, they only exist together with speech to emphasize, point out, animate or illustrate what is being said with words. Gestures are not typically expressing emotions, to the high degree that facial expressions do, but there are some gestures that for example express happiness, pride or anger and these gestures might be synchronized with speech. *Regulators* are being used to regulate the conversation. They, often head movements, are also speech dependent but they have nothing to do with the content of the words or the content of the conversation but are used only to regulate the speakers' and listeners' roles and management of turns. *Adaptors* especially refer to gestures that involve self-touching during speech but there are also other kinds of adaptors (touching/tapping objects like a pen). Usually they have nothing to do with the content of the speech or regulation of the conversation but they might still have an effect in some situations (Afifi, 2010).

Gibbs (1999) has pointed out that there seems to be a continuum of intentional and unintentional gestures. Intentional gestures are expressed because the communicator wants to share something or express something to others. Intentional also means that it is fully possible to inhibit these expressions. Unintentional gestures cannot be inhibited, at least not easily, and it is not even sure that the communicator is aware of the gestures produced. Intentional or not, most of these gestures are socially oriented. We don't use them when we are alone. Head nods used as feedback expressions are often used in a conversation but we are seldom aware of it. It means that they are intentional but on a low level of awareness and especially social oriented. Emblems are used intentionally but in this case we are aware of our use, and they also serve a social function. We don't produce affect displays when we are

alone. The question is, if we ever happen to talk to ourselves, do we use illustrators when we speak?

If we claim that “gesture and speech form a single, unified system” we will not just look at existing evidence but also try to find new evidence. To find new evidence we will try a new technique, a motion capture system, and a somewhat new test condition. First, the new technique is highly reliable and can produce large amounts of data to do trustworthy statistical analysis with. Second, the conditions that will be tested are based on participants that perceive themselves as bicultural and bilingual and that are asked to speak in one language in the first condition and in the other language in the second condition. If the participants have “a single, unified system” they will change their gestural pattern together with the spoken language.

The purpose of the present study, presented in this paper, is to find evidence for possible gestural patterns that are intertwined with each language used. To be able to do this we have to carry out several steps. The first step is to find existing empirical and theoretical evidence that possibly can support the integrated system model. We also have to create the test conditions that can help differentiate the alternative versions to an integrated model. One way to do this is to design a four-condition test environment. Two conditions are in one language and two are in another. Two conditions are in a face-to-face condition and two are in an audio only condition.

Table 1. Four different test conditions.

Type of interaction	Languages	
	Portuguese	Swedish
Face-to-face	Test condition 1	Test condition 2
Audio only	Test condition 3	Test condition 4

This makes four basic test environments that can be combined in four ways (see table 1). After a statistical analysis of the comparisons we will find out if the integrated systems hypothesis is supported, if any other hypothesis is supported or if

no hypothesis is supported. We will formulate two working hypothesis. The first one is the integrated systems hypothesis that can be compared to three other alternatives. The second working hypothesis is related to the languages and cultural backgrounds involved. Can we hypothesize that the gestural patterns are culture- and language relative? There is already some empirical support for this. We not only assume verbal code-switching (Bailey, 2010) when the participants change from one condition to another but also kinesic code-switching (cf. Burgoon, Buller & Woodall, 1996). If this hypothesis is supported it will also support the first hypothesis.

To find evidence to support that speech and gesture go together also in bilinguals we have tried to find cultures that are expected to differ to a high degree. If these conditions can't produce a culture-specific output to rule out the two system hypothesis then we have a weak case. We therefore looked for cultures that are very different on some cultural dimensions, cultures that have existed in different climates and that are based on different languages. Our best match for these criteria happened to be bicultural individuals that are both Swedish and Mozambican. In Sweden the main language is Swedish and in Mozambique the main (official) language is Portuguese. Sweden is a reserved, low contact and low context culture while Mozambique is an expressive, high contact and high context culture. If we can create a setting for these bicultural and bilingual individuals which, primed by the associated language, might elicit the related cultural gestural pattern then we can support the integrated communication systems hypothesis.

After a literature review we present our method in detail, we present the results of the four conditions and we discuss the outcome in the light of the working hypotheses and previous empirical and theoretical evidence.

One system or several communication systems?

This section will present some possible alternatives for how gesture and speech are related. If the primary belief is that speech and gesture are two communication systems that are tightly intertwined we first have to look at the alternatives. How plausible is the integrated hypothesis in the light of other

explanations? The alternatives are: two versions of a separated two systems model and two versions of a unified model.

The separated and independent systems. If we see communication production as based on two separate communication systems this can be understood in a strong sense and a weak sense. Two separated and independent systems that might co-occur but are not coordinated or synchronized is the strong version. The co-occurrence of speech and gesture is not grounded in the idea that one message might use both gesture and speech to be optimally delivered. The co-occurrence model is rather a suggestion that the speech system is delivering a deliberate message on a high level of awareness while the gesture system is delivering symptoms/signs of inner states that are existing on a low level of awareness. Since the two systems are separate there is no reason for the gesture system to adjust the output if the language produced by the speech system shifts from one to another.

The weaker side of the independent system hypothesis is what here will be called the auxiliary hypothesis. This position has been suggested by many but probably most strongly championed by Robert Krauss (1998; also Krauss, Chen & Gottesman, 2000). We use the gesture system as a support system when the verbal communication is especially complex or unusually difficult. Examples might be to find words or to express the meaning of a word or sentence. The gestures are often initiated before the production of the words as a way to activate the speech system in finding the words. When a short presentation is rehearsed, to take another example from Krauss studies, the frequency of gestures are lower than when a contribution is spontaneous. These examples are both in line with the idea that the gesture system is used as an auxiliary system to the speech system when it is needed.

The unified system(s). The unified systems alternatives can be formulated in a weak version and a strong version. The weak version is presupposing that we have two systems that are tightly intertwined to a degree that they are able to work as one system. The strong version is building on one single system that has two or more output channels/modalities.

The integrated systems hypothesis. The communication systems that work as being one system is the integrated model that originally is based on two systems that have been tightly integrated to work in a coordinated and synchronized way. This is the weaker version of a unified communication system. One of the most well-known advocates of this line is David McNeill (1992; 2007; McNeill & Duncan, 2000). McNeill's growth point theory is centered on the growth point that is a cognitive unit that, in its most minimal sense, lies behind the production of a message. A growth point is a unit for thinking-for-speaking. The unit grows into a potential complex expression involving both speech sounds and hand gestures. It should depend on the cognitive unit if both gesture and speech will be needed and used. According to Sowa et al. (2008) approximately 90 percent of the spoken utterances are accompanied by gestures. It means that only in some few cases the growth point is generating only speech and in even fewer instances are generating only gesture.

Hand to mouth hypothesis. The hand to mouth hypothesis is based on an evolutionary model that presupposed that the homo lineage (starting with homo erectus) were symbol minded before they could produce symbolic sounds with the speech organ but instead used gestures in a system that was like some form of sign-language. This hypothesis is suggested and supported by Michael Corballis (2002; 2007; also Gentilucci & Corballis, 2006; 2007) and Michael Tomasello (2008). The development started with indexical gestures like pointing to move on to iconic gestures and pantomime into symbolic gestures and further to resemble some simple version of today's signs language (Tomasello, 2008). Corballis (2007) is speculating about if the hand gestures partly changed to speech gestures. The movements made by the muscles in and around the speech organ are gestures that are necessary to produce speech sounds. The original hand gesture based communication system expanded to become also a mouth gesture system. Eventually the mouth gesture system became the dominant system (Gentilucci & Corballis, 2006). The original system has not developed into two systems but has just expanded to be able to process and produce more complex communication including words and hand

gestures instead of just hand gestures. This means that the intentional message production involves a choice of production means. If we need both gestures and words we produce messages based on both but if we need just one of the output channels/modalities we will produce messages based on only one of them. The one system hypothesis doesn't imply mandatory parallel production but optional output. An intentional output regulation means that we don't have to use gestures when we talk with someone on the phone and we don't have to speak or shout in an extremely noisy environment.

Empirical and theoretical evidence

Several studies have generated empirical support for one or two of the alternative communication systems above. These evidences will be presented to test if any of the hypotheses, especially the integrated, are probable. In some cases it is also possible to use other theories to support or explain tendencies or differences.

Empirical evidence. Self-synchrony is a phenomenon that William Condon in the 1960s observed on film recordings (Knapp & Hall, 2006). It seems as certain body parts are moving in synchrony with the words and clauses produced. These studies were followed up by Birdwhistell (1970) who called the synchronized body movements kinesic markers. The kinesic markers are, coordinated with the clauses, following a systematic pattern often related to spatial information or activities. Typical examples of kinesic markers are head movements (up and down), hand movements, eye lid movements (up and down) and eye brow movements (up and down). Kendon (1972) studied head and hand movements and found a similar pattern as previous studies. He concluded that the speech output and the kinesic output are two aspects of the same process. It should also be noted that gestures often accompany speech but it may also precede speech. All the mentioned studies support the unified system(s) approach. There is no other way to explain self-synchrony than to assume two tightly intertwined systems or one system with two output channels/modalities. The last observation about gestures preceding speech is more in line with the auxiliary approach.

Krauss (1998) and his colleagues carried out a number of studies to find out how speech is related to different kinds of hand gestures. They found that gestures were most often used together with words referring to activity (rather than passivity), words referring to concreteness (rather than abstractness) and words referring to spatiality (rather than non-spatiality). The most used gestures were directly related to spatiality. It was also found that gestures often preceded the spoken word, especially when it was related to spatiality. In another study they found that the participants slowed down their speech rate or did more speech errors when they were restricting the use of gestures and these differences typically occurred when the speech content was about spatiality. According to Krauss these results are supporting the idea that gestures are facilitators, that is, in line with the auxiliary hypothesis. This would also support the growth point theory since the spatial information that is about to be expressed might be a spatial mental image. The growth point needs both channels/modalities to be optimally expressed.

Butcher and Goldin-Meadow (2000; also Goldin-Meadow, 2003) describe the gesture use and development of children. During their first year children gesture sporadically and without accompanied words and they utter single words without any gestures. It is first when they become about two years of age that they start to use gestures together with words and that is about the same time as children start to combine words into two word sentences. This is in support of a two system model but it is also supporting a two system model that is becoming intertwined into a unified system.

Studies of deception and negotiation can be of value to the support of our alternative hypotheses. When a person is lying the rate of illustrators, that normally accompany speech, is reduced. An increase in self-adaptors is instead likely (Frank & Svetieva, 2013). In a separate two system model the self-adaptors are probable but the decrease in illustrators cannot easily be explained. From an auxiliary hypothesis perspective deception is a particularly complex and difficult situation. It would call for more gestures to support the deception, and the self-adaptors do not help.

Deception does not fit well with the auxiliary hypothesis. The integrated model is fundamentally a two system model that makes it possible to use the systems separately and the deception situation requires this but it doesn't work very well and that might be because the systems want to function in unity. The unified model is probably the easiest way to express something but lying is not in line with the optimal way of communication. The growth point theory is not making the explanation easier, though. The single system model, in the sense that it is intentionally driven, can shut down one of the channels/modalities for the purpose of deception. This is a possible explanation but it doesn't explain the increase of self-adaptors.

In negotiation situations, like bargaining, we sometimes produce incongruent messages. The words are saying "yes" to a proposal but the head or hands are saying no, or the words are saying "no" while the head or hands are saying yes or don't know. The latter outcome is likely when your bargaining threshold has been passed but when you still want more. The hands or head are signaling that the proposal is accepted but the words are saying that you are not satisfied yet (Boughton, 2013). The contradicting signals are possible for a separate two system model. The words are expressing what we are thinking and the gestures are expressing our emotions independent of each other. Contradicting signals are not supporting the auxiliary model, once again because the situation is complex and difficult and would call for supporting gestures instead of contradicting gestures. The integrated model can be in line with contradictory signals. Two ideas are expressed simultaneously but through different channels/modalities to express something whole nonetheless incongruent. This kind of situation is not easy to fit in to the growth point theory. To the single system model contradicting signals has to be some kind of paradox. How can one system express two contradicting messages at the same time? And why are these incongruent signs using one channel each? Why couldn't the single system deliberately be used for the intentional verbal message only?

Gestures are related to language and culture. If persons from an English speaking culture are asked to describe how a character is swinging in a rope they will use the word “swing” and do a swinging gesture but if persons from a Japanese speaking culture are asked to describe the same scene they have to express it with other words and other gestures because they don’t have an equivalent to the word “swing” (Kita, 2000). Since the word “swing” fits well together with one or a few gestures for the spatial swinging trajectory English speaking people need few gestures to retell the scene but the Japanese speakers both need more words to describe the scene and more gestures.

An older study by David Efron (see Kendon, 2004) uncovered in a detailed way how Italians and Jews gesture. There are obvious differences that can be directly linked to language and culture. When Italians gesture they move both the upper arm and the forearm in several directions. When Jews gesture they mainly move their forearm. Italians move the left and right arm in a symmetric way while the Jews move the right arm more than the left arm.

In the study about the swing the neat fit between word and gesture appears to be a support to the integrated systems hypothesis but the Japanese way to handle the lack of a suitable word is more in line with the auxiliary hypothesis. The Japanese have to gesture more because the complexity of the situation demands more gestures. On the other hand it can be said to depend on the characteristics of the language rather than the complexity. If language and gestures go together it can explain the different gestural pattern. This reasoning becomes even more obvious in the Efron case. Italians and Jews have different gesturing patterns that seem to be tightly connected both to the language and the culture. This difference is more in line with the unified system(s).

Several studies have shown that the perceptual system that we are equipped with is optimized for multimodal reception and interpretation. There are specialized areas in the brain that process multimodal information (Beauchamp, 2005). Meaningful symbolic sounds and symbolic gestures that are produced

simultaneously are processed in the same neural area even if the input was received through different sense modalities (Bernardis & Gentilucci, 2006). It has also been found that words and iconic gestures that are incongruent increase response time (time to process the information) and produce more errors in the receiver than words and gestures that are congruent (Kelly, Özyürek & Maris, 2009). It would be expected with a delay in time to process incongruent information if it is being processed as one unit instead of at least two. All these results are in support for a perceptual system that is prepared to process multimodal input, that is, both speech and gesture in a congruent and unified whole. We have this ability, it might be argued, because the production is integrated and produced as a unified message.

Theoretical evidence. Allwood (2008) is presenting a communication model that is based on three levels of awareness. On the lowest level we produce indexical signs. Most of them never become conscious to us. Communication on the first level is very fast, it is hard to control and we have low access to the process behind the produced signs as well as the produced signs themselves. On the second level, the mid-level, we process and produce iconic signs. We have higher access to both the process and the production of these iconic signs, we are able to control them a bit and the process as well as production is a bit slower than on the first level. The third level is the symbolic level. We have high access to the processing and production of symbols, we can control it to a high degree and the process/production is relatively slow. Speech is going on almost exclusively on the third level. Gestures exist on all levels. Indexical gestures are fast and uncontrollable, iconic gestures are a bit slower and relatively controllable and symbolic gestures are slow and highly controllable. If this model is plausible it can explain two aspects of human communication that is related to the present discussion. (1) We can deliberately choose to communicate emblems or not. We can decide if we want to express our thoughts or emotions with a symbolic gesture, a symbolic sound or a combination of both when the emblem and the word are congruent in meaning. (2) When we combine words with gestures on level one or level two it is most likely that the gesture will be produced before the

word(s). Indexical and iconic gestures are more quickly expressed. Spatial gestures are usually indexical or iconic. That might explain why spatial gestures precede spatial words when both are expressed.

Tomasellos (2008) communication theory is grounded in the human social cognition abilities. Humans are intentional beings that have goals and seek the right means to achieve their goals. The intention-reading ability that humans have is used to interpret the senders' intention with a message but it should be equally relevant to be able to read the receiver and foresee what the receiver will be able to understand or not. Tomasello is explaining that a person in a noisy environment realizes that the use of the voice is not enough and will therefore rely more on gestures. This is fully understandable if we see ourselves as intentional goal-oriented beings. The goal will be better achieved with the aid of gestures rather than with words that will drown in a sea of noise. With this said it would be equally reasonable to claim that a person will use less gestures in a dark room and less gestures while speaking on the phone. If it is a question of choosing the optimal means to attain the goal there is not much use of gestures while speaking on the phone. Tomasellos own theory would predict that the one system hypothesis will use separate output channels/modalities if that serves the goal. The growth point theory is based on another principle. It is not the goal that drives the output but the image or idea that is being expressed. This means that the one system hypothesis will predict markedly less gestures in a phone conversation compared to a face-to-face conversation while the integrated systems hypothesis only will have a limited decrease in gestures during a phone call.

A working hypothesis. All the empirical and theoretical evidence that is presented above is supporting the integrated hypothesis more than the other alternatives. Based on that we will formulate a working hypothesis:

H1. A unified communication system, based on two tightly integrated communication systems, will predict that a person that is bicultural and bilingual will change the gestural patterns depending on the language spoken. The obligatory tendency of the unified system will make individuals use gestures to a relatively high

degree also in an audio condition simply because it is easier to produce both than to shut down one system.

Affecting factors

There are two possible factors that can affect the test conditions. The first one is convergence/rapport building and the other one is priming.

Convergence and rapport building. Convergence, in accordance with the communication accommodation theory, is a way to express belonging and liking by the use of similar behavior as the co-communicator (Giles & Wadleigh, 2008; Giles & Soliz, 2015). This can be expressed as similar speech patterns (e.g. speech rate) or similar gestures. When individuals that interact feel a sense of mutual understanding, a harmonious relation, some kind of affinity or as if being on the same wavelength they are doing rapport building (Gibbs, 1999; van Meurs & Spencer-Oatey, 2010). Some of the most common ways to do this is to mirror each other (Matsumoto & Hwang, 2013c) or to do interactional synchrony (Egolf, 2012). Mirroring means that person A is producing a movement with a body part that is similar to what person B just did. The mirroring process is very fast and usually on a low level of awareness (Dimberg, & Thunberg, 1998; Iacobini, 2009; Goldman, 2013). Interactional synchrony is to produce synchronized movements that are either in synchrony with the speakers words (rhythm of the speech) or with the speakers body movements (Birdwhistell, 1970; Burgoon et al., 1996; Gill, 2008; Egolf, 2012). In an interview situation there is always a risk that the interviewers' movements are affecting the interviewees' movements. If the interviewee is the target of the study this affect might jeopardize the reliability of the result. It is important to be able to exclude convergence and rapport building.

Priming. There is a cognitive effect that is called priming. If a person is exposed by a certain cue (a word, an image or a movement) a memory, a whole memory system or a certain behavior may be triggered (Sobel, 2001; Baddeley, Eysenck & Anderson, 2009). Benet-Martínez, Leu, Lee, and Morris (2002) showed that Chinese-American biculturals displayed culturally congruent behavior when presented

with relevant cues associated with one of their cultural backgrounds. A culturally related cue can be an artefact or an image. It is also probable that a language that is associated with a culture has the function of a priming cue. An assumption is that when a bicultural individual receives the right cue he or she will shift from one cultural behavior to the one that is triggered by the cue. This means that it is not just the language that is shifted but also the typical, or a strong tendency towards the typical, behavior and bodily movement pattern that will be triggered to become displayed. Priming cues operate on a low level of awareness. It is thus hard for an individual to control all the effects of cue and memory system and the behavior that it triggers. To expect the use of a language to have a priming effect is connected to some uncertainty. Is the auditory stimulus enough or is it necessary to produce speech accompanied by associated bodily expressions and movements? This is very difficult to fully control for.

Cultural dimensions

Edward T. Hall (1969) suggested a cultural dimension that could describe differences in proximity, gazing patterns and touching patterns. People living in *contact cultures* like to stand close to each other when they are in a conversation. They like to stand in an angle that is letting them face each other, regular direct eye contact, speak with loud voices and to touch each other. People living in *non-contact cultures* prefer to stand at a distance, avoid regular and sustained eye contact and avoid touching (Martin & Nakayama, 2010). The terms used for the dimension has been changed a bit to *high contact cultures*, *mid contact cultures* and *low contact cultures* (Ting-Toomey & Chung, 2005). There are also very similar dimensions called immediacy orientation (Andersen, 1998), immediacy and expressiveness (Chen & Starosta, 1998) or expressive vs. reserved cultures (Matsumoto & Hwang, 2013b). The expressive vs. reserved dimension is including more channels/production modalities than the previous dimensions. It also focuses on variety, for example a variety of facial expressions and gestures used (cf. Young, 2011).

People living in cold Scandinavia are categorized as belonging to a low contact culture or a reserved culture (Ting-Toomey & Chung, 2005; Martin & Nakayama,

2010) while people living in warm regions belong to high contact cultures and expressive cultures (Lustig & Koester, 1993; Chen & Starosta, 1998). Based on this Sweden is a typical case of a low contact culture and a reserved culture. Mozambique is a high contact culture and an expressive culture (cf. Awa, 2009; Gesteland, 2012). We therefore hypothesize that:

H2. Bicultural and bilingual individuals that are partly Mozambican and partly Swedish will gesture more, with more intensity, when they talk Portuguese with a co-communicator talking the same language compared to when they talk Swedish with someone.

Methods

The selection, the procedure, reliability and how to analyze is described.

The selection

One member of the research group was herself bicultural and could speak several languages fluently. That affected our selection of the two cultural groups that the participants should belong to. We choose participants that had a Swedish-Mozambican bicultural belonging and thus could speak both Swedish and Portuguese. A specific bicultural identity combination was scoped out (one that included both a reserved and an expressive culture). Considering Benet-Martínez (2012) definition of multiculturalism as the experience of having been exposed to and having internalized two or more cultures - participants were chosen according to the following criteria:

1. They had to be fluent in both Swedish and Portuguese - the official languages of each culture.
2. They had to have spent at least 5 years in each country.
3. Had to still have connections in both cultures.

The participants. Based on these prerequisites, seven participants were chosen. Out of these seven participants, four were female and three were male. Six were born in Mozambique and had moved to Sweden later in life. All seven were either children of or spouses in mixed marriages between Swedish and Mozambican

partners. The average age was 33, with the oldest participant at 61 and the youngest at 19.

Procedure

The setting. For the purpose of this paper, participants were interviewed one at a time while being filmed using Motion Capture technique which focused on their upper body movements. Practically, this entailed for participants and the interviewer to be filmed while wearing sensors on key-parts of their upper body which would later be used to measure expressiveness versus reservation. For this study, eight oqus cameras were used calibrated before each interview-session to the QTM software (see below). All interviews were carried out in the same studio, equipped with motion capture cameras and software, taking about one hour to carry out each and rendering approximately 40 minutes interview-time with each interviewee dependent on the length of time taken in each interview. Average time for each interview block = 10 minutes.

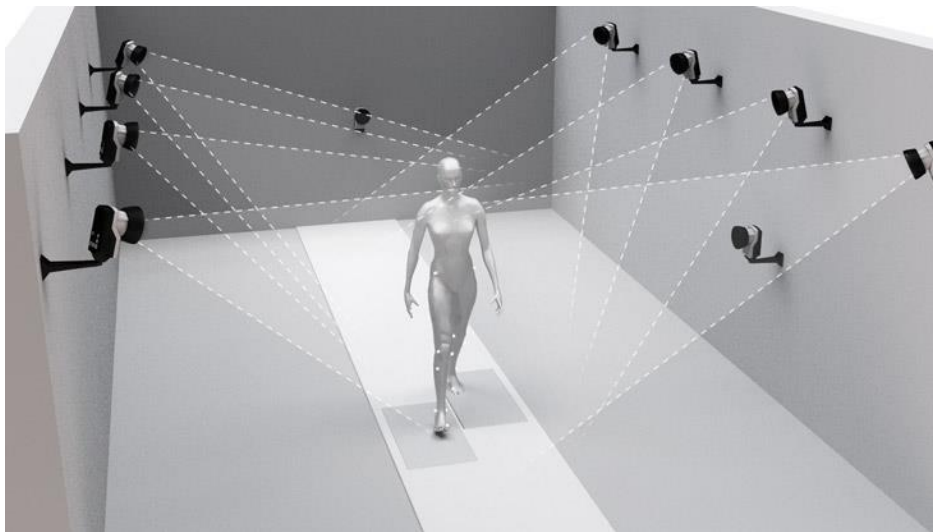


Figure 1. Motion capture studio equipped with motion capture cameras. (taken from Qualisys.com)

The software used. Qualisys Track Manager, hereby called QTM, is a motion capture software used for tracking movements by filming sensors through oqus and pro reflex cameras connected together. Oqus cameras are designed to capture

accurate mocap data, meaning that it is capable of calculating marker positions with accuracy and speed, allowing for the documentation of hundreds of markers during thousands of frames per second – run from an ordinary laptop with the correct software. The software, QTM can then join up the data (x, y) collected in 2D from various cameras into 3D (x, y, z) positions. QTM was used for this study due to its ability to track motion in real time and high accuracy in capturing each marker (Qualisys.com). The software can produce different kind of output. We used velocity and it means that the output value is in millimeters per second. Velocity, thus, gives the speed of a marker. Higher velocity means a higher degree of movement. It can also be understood as high intensity.

Markers. In order to be able to capture the movements of body parts through motion capture filming using QTM, sensors were placed on specific body parts of both interviewee and interviewer for each interview. Each area marked with a sensor is referred to as a marker, and there were a total of 21 markers on each participant during the study, on the following points:

Forehead

Left ear

Right ear

Neck

Left shoulder

Right shoulder

Left hip

Right hip

Tailbone

Left elbow

Right elbow

Left hand:

Inner left wrist

Outer left wrist

Left pinkie finger knuckle

Left index finger knuckle

Left index finger

Right hand:

Inner right wrist

Outer right wrist

Right pinkie finger knuckle

Right index finger knuckle

Right index finger

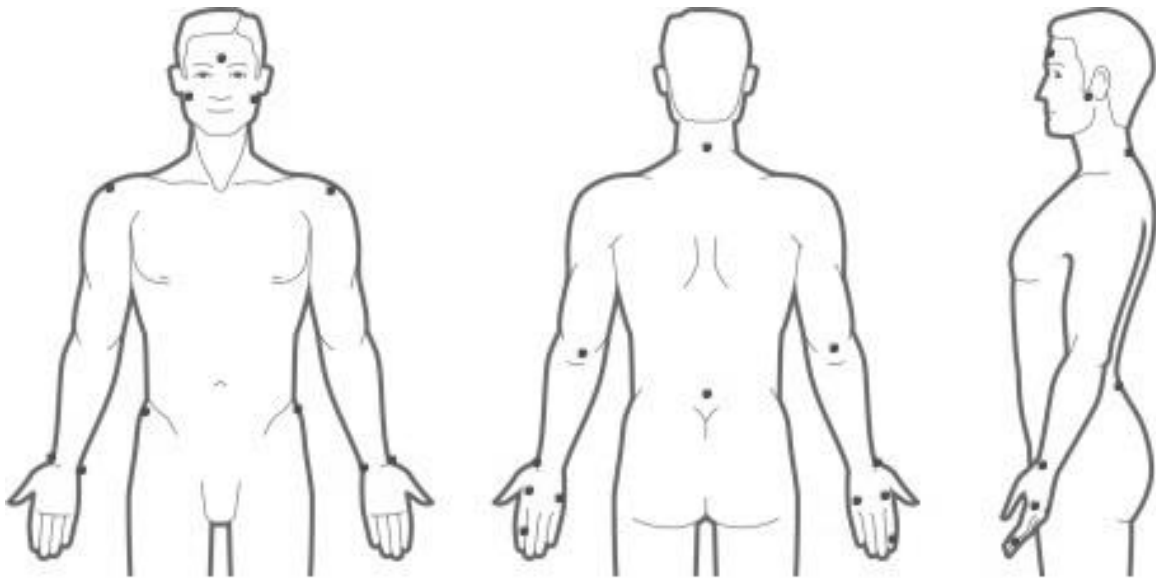


Figure 2. Markers used during interviewing

Interviewing. Data were collected during semi-structured personal interviews. An Interview Guide Approach allowed us to explore specific topics and build a conversation with the participants. Using this approach provides a degree of flexibility and facilitates an elaboration of participants' experiences and interpretations (Hennink, Hutter & Bailey 2011). The goal of each interview is firstly to allow the participant an opportunity to talk in the language of the interview, priming the related cultural frame, secondly, to let them relax and talk freely, in order to minimize the effects of nervousness related to being interviewed or documented.

The interview was divided into four-block stages where two blocks were done face-to-face and two blocks were done through Skype with only the auditory communication channel in use. The average interview-time for the face-to-face blocks was 10 minutes each and for the audio-only blocks it was 9 minutes each. The blocks are described below:

Block 1: Face-to-face interview in Portuguese

Block 2: Face-to-face interview in Swedish

Block 3: Skype (audio) interview in Portuguese

Block 4: Skype (audio) interview in Swedish

These blocks were done in alternating order for each interviewee in order to avoid possible patterns followed by maintaining a specific order. This also prevented habituation and memorizing from one language to another in a systematic way.

Semi-structured interviewing was here chosen as the most appropriate interaction method to prime cultural frame switching, as it is “primarily used when you seek to capture people’s individual voices and stories.” (Hennink et al., 2011, p. 110) and it also means that both parties speak recurrently (in a structured interview the interviewer would speak more and in a unstructured interview the interviewee would speak much more which might reduce the priming effect). Considering the goal of our interviews, focus was more on asking more open-ended questions (Hargie, 2011) in order to allow participants the chance to talk about themselves and engage their cultural frames through this process. Following Goldin-Meadow’s theory on a unified system between the verbal and nonverbal gestural communication channels, the interview’s focus was on creating an environment where speech would be allowed to influence the participants’ nonverbal behavior freely. For example, instead of asking - Do you define yourself as bicultural? We asked - Could you describe your cultural identity? The interview guide for block 1 and block 2 were the same but in different languages and the interview guide for block 3 and block 4 were the same but in different languages. Even if the questions

differed between the first two blocks and the next two blocks the theme of the questions were the same: living as a bicultural person.

Reliability

In order to look at reliability, the degree to which an assessment tool produces stable and consistent results, it is necessary to look firstly at the main tool used for documenting the experiments, namely the QTM-software. It's high-end, high-precision quality has allowed this to be a software used not only in media and entertainments, but also for documenting biomechanics and industrial applications. This software is quite user-friendly, yet requires one to keep in mind visibility issues that may occur during documentation. These can occur when markers are very close together and moving a lot, thereby obscuring each other for a few frames or if something else obscures vision during documentation. A practical example of this would be while capturing the individual's hand-markers, if the individual crosses his/her arms with the hands underneath, making it hard for the cameras to document this. When visibility is not an issue, a marker will be captured all in one continuous trajectory throughout the interview allowing for a 100% documentation of the marker. In some cases, when visibility does become an issue, the trajectory will be divided into when the marker was last seen and the next time it is seen – so the trajectory is divided into parts. In this study, an Automatic Identification of Markers or AIM-model was created in order to allow the model to automatically identify all the obvious trajectories. The remaining trajectories that were then in parts had to be manually identified at a later stage. This allowed for a higher reliability-level of the QTM-program's marker identification. Figure 3 below illustrates what the markers look like once they have all been identified in the QTM-program.

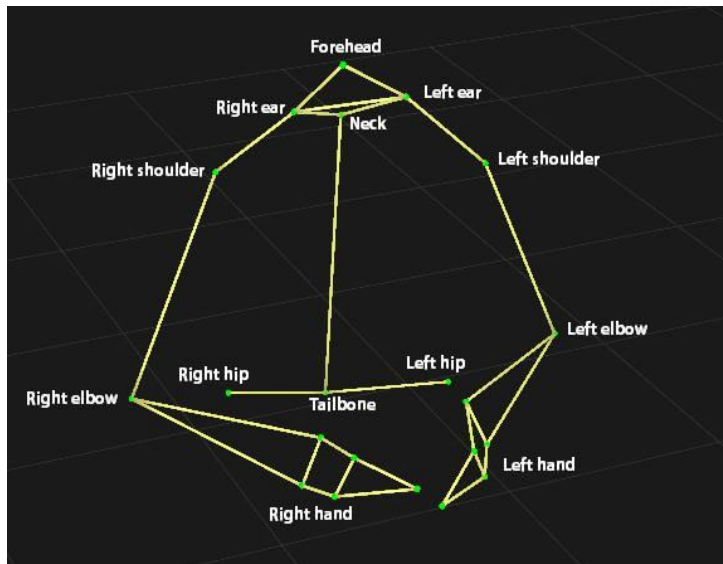


Figure 3. *Snapshot of identified markers on a participants upper body. The forehead marker on top.*

The forehead marker had a 100% visibility and therefore a 100% reliability. The relatively low number of participants is not a reliability problem nor a validity problem since the QTM system is producing huge amounts of data for each individual (see below). The purpose is to find out if the participants shift gestural patterns when they switch cultural identity/language. This is exactly what we can find out with the used system. Generalizability on the other hand is nothing that we are speculating or make claims about.

How to analyze the data

The QTM software produces 170 frames per second (it can produce up to 300 frames per second) in the present study¹. Each frame gives a value. That means that each marker produces 170×600 data units per interview summing up to approximately 102 000 data units per marker. To calculate the mean for all participants forehead movement in the face-to-face conditions we have to add 102 000 seven times for the Swedish block and seven times for the Portuguese block. It will produce the average of 1 428 000 data units in the face-to-face blocks and

¹ Since the cameras used can record 1000 frames per second the frames that the QTM is producing is an average of the camera frames that are merged into the QTM frames.

nearly the same number of data in the skype/audio blocks. These huge amounts of data for just one kind of marker are almost too much to handle. We solved this by calculating an average per minute for each marker and used these new values to do t-tests on the average of different blocks or conditions and correlation tests. If the differences were not large enough to be significant we used the full data set. It should be understood that when the full data set is used to do t-test even very small differences become significant. This is another reason why we avoided the full data set. We wanted to focus on more obvious differences if they could be found.

Arms and hands moved most, something that was expected. Instead of calculating all of the markers on arms and hands we selected the two that moved the most: the left and right index finger. The third body part that moved much was the head. We selected the forehead marker, since it moved most of the head markers. In a few calculations we also used the ear markers, both left and right ear. The three main markers were statistically analyzed both separately and together.

First we did comparisons between the velocity of the selected markers in block 1 and block 2. The average of all left index fingers in block 1 was compared with all left index fingers in block 2 and the difference between the mean was t-tested. The procedure was repeated for the right index finger and the forehead. Within blocks we tested for correlations between the markers to find unique cultural patterns. These calculations were used in all comparisons between blocks and conditions.

Results

The first test is to compare face-to-face conversations with audio conversations. The average velocity of the left index finger in block 1 and 2 (face-to-face) is 128,36 mm per second while the average velocity in block 3 and 4 (audio only) is 92,02 mm per second. The velocity in the audio condition is reduced with more than 28 percent (the difference is significant, $p < 0,01$) compared to the face-to-face condition. The average velocity for the right index finger is higher, 148,92 mm per second in block 1 and 2, and is reduced with almost 26 percent to 110,44 mm per second in block 3 and 4 (the difference is significant, $p < 0,01$). When both left and right index fingers are

combined in block 1 and 2 and compared with both markers in block 3 and 4 the difference is 27 percent ($p < 0,001$). The intensity of the hand movements is distinctively lower in the audio condition. A simple explanation is that the hands are used less when the communicators can't see each other. This does not suggest what kind of gestures that are reduced. Emblems lose their function if they can't be seen so maybe they have disappeared. With this line of argumentation it might be reasonable to ask why other types of gestures haven't fully disappeared. The receiver can't see any of them. Another suggestion is that adaptors are reduced in the audio condition. Individuals that feel worried or stressed when they are observed might be less affected in an audio condition. Gestures used for turn management are maybe the best candidate since they are not needed to regulate the conversation in the audio condition. There are still a lot of hand movements in the audio condition, only 27 percent lower than in the face-to-face condition. What is the function of these movements? Are they illustrators that unconsciously accompany speech and/or self-synchrony movements? One thing can be stated with a high certainty and it is that the hand movements in the audio condition are too frequently occurring to be produced by a communication system that is not integrated with the speech system. If the gestures alone have no function they should be almost zero. The only movements that can have any reasonable function are self-adaptors, for example if some part of the body is itching. It is very easy to see in the recordings that there are few adaptors in general compared to hand movements in front of the body.

Head movements have also decreased in velocity when comparing face-to-face with audio conditions. The average velocity is 56,95 mm per second in block 1 and 2, and 50,60 mm per second in block 3 and 4. The velocity is reduced with more than 11 percent (the difference is significant, $p < 0,01$). With a help from the ear markers it is possible to differentiate between head movements on a horizontal plane (movements side to side like headshakes) from head movements on a vertical plane (movements up and down like head nods). The horizontal movements do not decrease at all between the conditions. It is only the vertical head movements that have been reduce

from the face-to-face condition to the audio condition. Vertical head movements like head nods are used in turn management (Duncan, 1974) and especially as feedback (Allwood & Cerrato, 2003; Afifi, 2010). These kinds of movements seem to have been reduced in the audio conditions since they don't serve its purpose anymore. The remaining head movements probably primarily serve a self-synchronization purpose (cf. Knapp & Hall, 2006). Head movements may also be integrated with speech. It is, also in this case, difficult to explain the function of all head movements if it is not related to speech since they are not serving any function outside of the conversation.

All standard deviations are lower in the audio condition, all except one marker. The head marker in block 4 did not decrease. High standard deviation is suggesting that the velocity is varying a lot from the average. There is more variation in speed during the face-to-face condition compared to the audio condition. This might suggest that the level of expressiveness is higher in the face-to-face condition. Why, because the speech is more varied? This was not recorded in the present study.

The next step is now to compare language blocks. Block 1 and 3 are performed in Portuguese while block 2 and 4 are performed in Swedish. When all three markers in block 1 and 3 were compared with all three markers in block 2 and 4 the velocity in the latter blocks have decreased with more than 6 percent (the difference is significant, $p < 0,05$). Both hand movements and head movements have a higher velocity when the participants are speaking Portuguese compared to when they are talking Swedish. Also in this comparison the standard deviation is higher in the Portuguese blocks. The variation in the gesture velocity is higher when the participants speak Portuguese. Preliminary this concludes that the speech communication system has to be integrated with the gesture communication system. There is no other strong explanation why the gesture patterns change when the language used change.

To find more details of value it is time to look closer at the two face-to-face blocks in a comparison and also look at the two audio blocks in a comparison. There are more differences between block 1 and 2 than has been revealed about the other

comparisons so far. The two index fingers combined in block 1 have an 18 percent higher velocity ($p < 0,05$) than the same markers in block 2. The head markers in block 1 have an almost 10 percent higher velocity ($p < 0,05$) than the same marker in block 2. Just as in the overall differences between languages the three markers have a higher intensity when the participants speak Portuguese. The standard deviation is also higher in block 1 which means that the variation in gesture velocity is higher when the participants speak Portuguese.

Generally the left hand and left index finger has a lower velocity compared to the right hand and right index finger. This difference is more pronounced when the participants are talking Portuguese compared to when they are talking Swedish. The left index finger has a 17 percent lower velocity than the right index finger in block 1. The left index finger has a 9 percent lower velocity than the right index finger in block 2. Both differences are statistically significant ($p < 0,01$). It is obvious though that the asymmetry is larger in the Portuguese condition. The right hand is much more active and intense.

The head movements in a horizontal as well as a vertical plane differ between the two blocks. In block 1 when the participants speak Portuguese they move their heads 14 percent more in the horizontal plane compared to block 2. In block 1 they also move their heads 6 percent more in the vertical plane. The general tendency is to move the head more while speaking Portuguese but the most striking difference is that there is a pattern for head movements side to side (e.g. headshakes).

The three markers in each block can be tested for correlation. In block 1 there is a very high correlation in velocity between the left index finger and the right index finger ($r = 0,95$). The finger movements also correlate with the head movements. The left index finger has a rather high correlation to the forehead ($r = 0,63$) and the right index finger has an almost as high correlation to the forehead ($r = 0,61$). This can be compared to block 2 when the participants talk Swedish. The correlation between the left index finger and the right index finger is very high ($r = 0,93$), again. The correlation between the left index finger and the forehead is lower than in block 1

($r=0,48$) and the same tendency is found between the right index finger and the forehead ($r=0,51$). All of these correlations are highly significant. The interesting aspect is the higher coordination and synchronization that is performed when the participants speak Portuguese. The body movements are more in concert with each other when the participants speak Portuguese.

All in all the face-to-face condition is strongly indicating that the use of Portuguese is combined with a specific gestural pattern that includes higher velocity in general, higher variety in the velocity, a stronger asymmetry between the right hand and the left hand, more horizontal head movements and higher correlations (more coordination and synchronization) between hand and head movements.

To make sure that all these effects aren't just mirroring behaviors by the participants in relation to the interviewer we have to take a closer look at block 3 and 4. First it can be stated that the asymmetry between right and left hand while speaking Portuguese can't be a mirroring behavior since the interviewer didn't have the same tendency. In a comparison between block 3 and 4 the simplified data set isn't good enough. We have to use the full data set.

The velocity of the left index finger is 8,5 percent higher in block 3 compared to block 4 ($p<0,001$). The velocity of the right index finger is 9 percent higher in block 3 compared to block 4 ($p<0,001$). Differences in head movements are just above 1 percent, that is a slightly higher velocity in block 3, but the difference is significant ($p<0,001$). The velocity is less than 1 mm per second lower in the Swedish condition. Perfectly in line with this tendency is the standard deviation. It is higher in block 3 for both fingers but not for the head.

The asymmetry between right hand and left hand is now stronger in the Swedish condition, compared to the face-to-face condition, but there is still a higher asymmetry in the Portuguese block. Even if the head movements don't differ in velocity they still differ when it comes to horizontal movements. The participants more often move their head sideways (e.g. headshakes) when they speak Portuguese compared to when they speak Swedish. Hand and head correlations are generally

lower in the audio condition compared to the face-to-face condition but the most obvious tendency is that the correlations in the Swedish audio condition has decreased. The lowest correlation that is to be found between the right hand and the head is much lower but still significant ($r=0,28$; $p<0,05$). The correlations in the Portuguese condition are ($r=$) 0,50 or higher.

Table 2. Average gesture velocity in different conditions.

	Face-to-face			Audio		
	Left index finger	Right index finger	Forehead	Left index finger	Right index finger	Forehead
Portuguese	138 mm/s.	166,9 mm/s.	59,9 mm/s.	94,3 mm/s.	114,2 mm/s.	51 mm/s.
Swedish	118,7 mm/s.	130,9 mm/s.	54 mm/s.	89,8 mm/s.	106,7 mm/s.	50,2 mm/s.
Summary	128,4 mm/s.	148,9 mm/s.	57 mm/s.	92 mm/s.	110,5 mm/s.	50,6 mm/s.

To sum the last comparison, the tendencies from the face-to-face condition is somewhat weakened in the audio condition. There are still significant differences in the hand velocity and strong differences in the correlations between hand and head movements. The asymmetry and the horizontal head movements are reduced and the head movement velocity is almost vanished.

Discussion

The present study is centered around two working hypotheses. They can both be supported independent of each other but if the second hypothesis is verified it will give a strong support to the first hypothesis. We therefore start to look closer at the second hypothesis. In a comparison between the two Portuguese blocks and the two Swedish blocks there is immediate support for the second hypothesis. All three markers have a higher velocity in the Portuguese condition compared to the Swedish condition (the differences are significant). There is also a higher variation in the Portuguese condition. We can already discern two different cultural gesture patterns. In the face condition even more tendencies become revealed. Except for the higher velocity in hand and head movements and the higher variation in velocity during the

Portuguese conversation there is also a higher asymmetry between the right and left hands, more head movements on the horizontal plane and higher correlations between the three body parts. The Mozambican cultural gesture pattern seems to be more intense, more varied and more internally coordinated/synchronized. It fits well with the expressive style (Matsumoto & Hwang, 2013; Young, 2011). The conversations in Swedish are accompanied by a less intense, less varied and less coordinated/synchronized gesture pattern. The participants behave more reserved when they speak Swedish or at least less expressive.

To be fully sure about this difference between cultures and languages we also tested the audio condition. The differences between Portuguese and Swedish decreased a bit and in some aspects it was almost gone. The differences that definitely remained were between hand movements, between variation and between correlations. It is still safe to claim that hand movements are more intense and more varied and that the coordination between hands and head is higher when the participants speak Portuguese. The almost diminished asymmetry between hand movements and the almost vanished difference between head movements can maybe be explained. We tested for correlations in the face-to-face conditions between the interviewer and the interviewees and the interviewer constantly had a lower intensity on all markers that were compared. The highest correlation was found between the interviewers head and shoulder markers and the same markers on the interviewees during the Swedish conversations. Since the study took place in a Swedish environment and all except one participant (including the interviewer) was born in Mozambique it would be expected to find the highest correlations in the Portuguese conditions because of a sense of identification with a fellow countryman. The closest mirroring or interactive synchrony was not found in the Portuguese block. The asymmetry between left and right hands during the Portuguese condition was not caused by the interviewer since the interviewer did not display that asymmetric pattern. Even if there might have been mirroring and interactive synchrony during the face-to-face conditions it did not seem to cause a systematic cultural pattern in

gestural behavior. The pattern that can be seen in the face-to-face condition is probably rather a behavior that can be expected during face-to-face interaction, the participants became more Mozambican when they got to talk Portuguese face-to-face with someone that could speak in a Mozambican way. That might also have included the corresponding body movements. We cannot be sure that the priming cue is the language in itself but maybe the language and the gestures in coordination. Part of those characteristics disappears in an audio condition. The cultural dimension is partly reduced when they only hear each other. The priming component is weaker. One reason that the head movements were more reduced in the Portuguese block in the audio condition compared to the decrease in the Swedish condition might be that head movements are part of the Swedish signature. Swedes do not move their arms a lot but they have relatively intense head movements, especially vertical movements (like head nods) (Gesteland, 2012). The participants stayed true to the Swedish gestural pattern also during the audio conversation (block 4) and they also seemed consistent with the Portuguese pattern when they kept moving their heads more on the horizontal plane (headshakes) in block 3.

Even if the last differences mentioned are not strong indicators of cultural differences the whole picture is pretty convincing. The bicultural and bilingual participants changed gestural pattern when they talked Portuguese compared to when they talked Swedish. Hypothesis 2 is supported and verified.

If the test conditions are to be in line with the separate two system model there should be no systematic differences between the language conditions. Since there are differences it is a drawback for that model. There should, on the other hand, be differences between the face-to-face condition and the audio condition if it can't be proved that the gestures in the audio condition mainly are of an indexical kind and irrelevant for speech. The claim is not convincing. It is also less convincing because of the higher level of correlation that was found between hands and head in the face-to-face condition compared to the audio condition. This is very difficult to explain with a separate two system model. Why should the coordination and synchronization be

higher face-to-face? It is almost as difficult to explain why the variety in velocity is higher in the face-to-face condition compared to the audio condition. The support for the separate two system model is low.

The auxiliary model is suggesting that the gestures will increase if the content is complex and difficult. Since the order of the four blocks was altered for each participant we can't expect any block to systematically be more difficult than the others. Also, since the content of the interviews were all about rather abstract topics like cultural identity it should both be a high level of complexity (and therefore rather difficult) but also far from concrete and spatial. We cannot assume that the face-to-face condition is more difficult or more concrete/spatial than the audio conditions. The face-to-face conditions are the ones that generated more gestures, more intense gestures. We cannot either claim that the Portuguese conditions are more difficult or more concrete/spatial than the Swedish conditions. The auxiliary hypothesis will not get support from this study.

A basic assumption about the one system hypothesis is that we don't have to learn to use both speech and gesture (but this assumption is not supported, Butcher & Goldin-Meadow, 2000). We have to learn to use only one of them separately and that learning is based on deliberate use to achieve communicative goals. Just as we have learned that there is no motivation to use speech in a noisy environment there is no motivation to use gestures in a dark room. Even if the second hypothesis is verified and that also is a support for the one system model it is not easy to explain why the participants use many and relatively intense gestures when they communicate via audio only. It doesn't make it easier to attain the communicative goals. It is possible for the one system model to explain why there is a higher variance in the Portuguese condition compared to the Swedish and why there is a higher correlation between hands and head in the Portuguese condition but it is more difficult to explain why there is more variance in velocity and higher correlation in the face-to-face condition compared to the audio condition.

There are a lot of abilities that we have learned that become stronger than the innate tendencies. (1) We have learned to focus on written text that has no attraction on attention compared to moving things, bright lights, sounds, faces and so on. All these other kinds of stimuli are innately something that our attention system is drawn to but we anyway have the ability to focus on lifeless black figures on lifeless white paper (or on a screen) also when there is other kind of stimuli around. Learning has created a strong tendency (Proulx, 2007; Young, 2011). (2) We have learned to combine voice and lip movement when we process and interpret speech. Children can use them separately when they determine what they hear. Adults can't help to hear a fused or combined sound (McGurk & MacDonald, 1976). Once we have learned, it overrides the separate perceptual systems. With a similar argument we can say that the integrated two system hypothesis, that is a communication system functioning as one system, is based on learning that has made it easy to use both systems together and instead demands extra effort to use the systems separately. Children don't use both systems in an integrated way (Butcher & Goldin-Meadow, 2000) but in a few years they learn to do that. Learning has a strong influence also in this case.

In a comparison between the face-to-face condition and the audio condition it can be assumed that the gesture velocity can be high in both conditions. The main reasons for this are an unaffected self-synchrony that has a similar function in both conditions and a high level in the use of illustrators. Since the level of gesture velocity is relatively high in the audio condition we can assume that it basically is caused by self-synchrony and the use of illustrators. The decrease has to be explained as well: (1) the use of emblems is not a mandatory tendency and therefore that category of gestures may have been reduced. We can deliberately inhibit emblems or not produce them in the first place just as we can chose what words to say and not since they are all produced on a high level of awareness (Gibbs, 1999; Allwood, 2008); (2) the use of regulators is strongly reduced since there is no face-to-face interaction to regulate. This is especially true for the head nods (Afifi, 2010) that obviously were

reduced in the audio condition. Even if this is normally going on on a low level of awareness the context affects us to use this category of gestures less in audio conditions; (3) the interactive synchrony has no function in audio conditions; (4) the tension is probably lower for all participants in the audio condition when no one is observing them and that will have made them less nervous, something that probably will have reduced the self-adaptors. With these four reasons and maybe some more it is a bit surprising that the level of gesture velocity is still relatively high in the audio condition. It can't easily be explained in other ways than a more or less automatic and mandatory use of two intertwined systems.

The integrated system hypothesis is the only alternative that can explain why the participants are gesturing with a higher variety in some conditions, with higher correlations between hands and head in some conditions and with an asymmetry between left and right hands in some conditions. It is simply because speech and gesture go together in a primed cultural context. This is what the individuals have learned when they have been exposed to others communicative behavior.

Previous empirical evidence and the evidence from the present study are partly supporting the one system hypothesis but are most strongly supporting the integrated systems hypothesis. All in all the final step has to be to verify the first working hypothesis. This study's empirical tendencies are most strongly supporting the integrated systems hypothesis. The results do not fit the other hypothesis equally well.

The special case of bilingual participants might confuse the understanding about language and the integrated systems hypothesis. What is the relation between a language and a unified speech system? The simple answer is that the speech system is producing the language, it makes it go from an internal mental structure to an external shared behavior. If a person has two internal mental language systems they will both use and rely on the single speech system. Brocas aphasia is a good proof for this relation. If a bilingual person is involved in an accident and the part of the brain that is responsible for producing speech, Brocas area, is damaged that person will

lose his or her ability to speak fluently in both languages (Fabbro, 2001). This is of course also what will happen to a monolingual person. That is why we have to differentiate language from speech production. According to McNeill (1992) aphasia is affecting both the speech and gesture ability. This is another support for the integrated systems hypothesis but this kind of claim calls for caution. There are many types of aphasia and all of them don't include impairment in the gesture ability (Ahlsén, 2008). This might be a way to say that there is a lot of support for the integrated systems hypothesis but there are also many uncertain factors that still can't be explained.

Conclusions

The purpose of the study is to find support for the integrated systems hypothesis. Two working hypothesis were formulated. The second, and secondary, hypothesis was verified: the bicultural and bilingual participants change their gestural pattern depending on what language they speak. The tendency is somewhat stronger in the face-to-face condition. The first, and primary, hypothesis was supported: the only hypothesis that can fit all the results in this study, all four test conditions, is the integrated systems hypothesis. The participants most likely use two intertwined communication systems when they communicate in an interpersonal situation.

References

- [1] Afifi, W. A. (2010). Nonverbal communication. In Whaley, B. B. & Samter, W. (eds.) *Explaining communication. Contemporary theories and exemplars*. London: Routledge.
- [2] Ahlsén, E. (2008). Neurological disorders of embodied communication. In Wachmuth, I., Lenzen, M. & Knoblich, G. (eds.) *Embodied Communication in Humans and Machines*. Oxford: Oxford University Press.
- [3] Allwood, J. (2008). A Typology of Embodied Communication. In Wachmuth, I., Lenzen, M. & Knoblich, G. (eds.) *Embodied Communication in Humans and Machines*. Oxford: Oxford University Press.
- [4] Allwood, J., & Cerrato, L. (2003). A study of gestural feedback expressions. In The First Nordic Symposium on Multimodal Communication. Copenhagen.
- [5] Awa, N. E. (2009). Communication in Africa: Implications for development planning. *Howard Journal of Communications*, 1 (3), pp.131-144.
- [6] Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2009). *Memory*. New York: Psychology Press.
- [7] Bailey, B. (2010). Multilingual forms of talk and identity work. In Matsumoto, D. (ed.) *APA handbook of intercultural communication*. Washington, DC: American Psychological Association.
- [8] Beauchamp, M. S. (2005). See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Current Opinion in Neurobiology*, 15, 145–153.
- [9] Benet-Martínez, V. (2012). Multiculturalism: Cultural, Social, and Personality Processes. In Deaux, K. & Snyder, M. (eds) *Oxford handbook of personality and social psychology*. Oxford: Oxford University Press.
- [10] Benet-Martínez, V., Leu, J., Lee, F., & Morris, M. (2002). Negotiating Biculturalism: Cultural Frame Switching in Biculturals with Oppositional Versus Compatible Cultural Identities. *Journal of Cross-Cultural Psychology*, 33 (5), 492–516.
- [11] Bernardis, P., & Gentilucci, M. (2006). Speech and gesture share the same communication system. *Neuropsychologia*, 44, 178–190.
- [12] Birdwhistell, R. L. (1970). *Kinesics and context*. Philadelphia: University of Pennsylvania Press.
- [13] Boughton, A. (2013). Negotiation and nonverbal communication. In D. Matsumoto, M. G. Frank & H. S. Hwang (eds.), *Nonverbal communication. Science and applications*. London: Sage.
- [14] Burgoon, J. K., Buller, D. B., & Woodall, W. G. (1996). *Nonverbal communication. The unspoken dialogue*. New York: MacGraw-Hill.
- [15] Butcher, C., & Goldin-Meadow, S. (2000). Gesture and the transition from one- to two-word speech: when hand and mouth come together. In McNeill, D. (ed.) *Language and gesture*. Cambridge: Cambridge University Press.

- [16] Callan, D. E. et al. (2003). Neural processes underlying perceptual enhancement by visual speech gestures. *NeuroReport*, 14 (17), 2213–2218.
- [17] Chen, G-M., & Starosta, W. J. (1998). *Foundations of intercultural communication*. Boston: Allyn and Bacon.
- [18] Corballis, M.C. (2002). *From Hand to Mouth: The Origins of Language*. Princeton, NJ: Princeton University Press.
- [19] Corballis, M.C. (2007). The evolution of language. In Fiedler, K. (ed.) *Social communication*. New York: Psychology Press.
- [20] Dimberg, U., & Thunberg, M. (1998). Rapid facial reactions to emotional facial expressions. *Scandinavian Journal of Psychology*, 39 (1), 39–45.
- [21] Duncan, S. Jr. (1974). On the structure of speaker-auditor interaction during speaking turns. *Language in Society*, 2, 161–180.
- [22] Egolf, D. B. (2012). *Human communication and the brain*. New York: Lexington Books.
- [23] Ekman, P. & Friesen, W. V. (1981). The repertoire of nonverbal behavior: categories, origins, usage, and coding. First published in 1967. In Kendon, A. (ed.) *Nonverbal communication, interaction, and gesture*. The Hague: Mouton Publishers.
- [24] Fabbro, F. (2001). The bilingual brain: bilingual aphasia. *Brain and Language*, 79, 201–210.
- [25] Frank, M. G., Maroulis, A., & Griffin, D. J. (2013). The voice. In D. Matsumoto, M. G. Frank & H. S. Hwang (eds.), *Nonverbal communication. Science and applications*. London: Sage.
- [26] Frank, M. G., & Svetieva, E. (2013). Deception. In D. Matsumoto, M. G. Frank & H. S. Hwang (eds.), *Nonverbal communication. Science and applications*. London: Sage.
- [27] Gentilucci, M., & Corballis, M. (2006). From manual gesture to speech: a gradual transition. *Neuroscience and Biobehavioral reviews*, 30, 949–960.
- [28] Gentilucci, M., & Corballis, M. (2007). The hominid that talked. In Pasternak, C. (ed.) *What makes us human?* Oxford: A oneworld book.
- [29] Gesteland, R. (2012). *Cross-cultural business behaviour*. Copenhagen: Copenhagen Business School Press.
- [30] Gibbs, R. W. (1999). *Intentions in the experience of meaning*. Cambridge: Cambridge University Press.
- [31] Giles, H., & Wadleigh, P. M. (2008). Accommodating nonverbally. In Guerrero, L. K. & Hecht, M. L. (eds.) *The nonverbal communication reader*. Long Grove, IL: Waveland Press.
- [32] Giles, H., & Soliz, J. (2015). Communication accommodation theory: a situated framework for relational, family, and intergroup dynamics. In Brathwaite, D. O. & Schrodtt, P. (eds.) *Engaging theories in interpersonal communication. Multiple perspectives*. Los Angeles: SAGE.

- [33] Gill, S. P. (2008). Knowledge as embodied performance. In Gill, S. P. (ed.) *Cognition, communication and interaction*. London: Springer.
- [34] Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge: Harvard University Press.
- [35] Goldman, A. I. (2013). *Joint ventures*. Oxford: Oxford University Press.
- [36] Hall, E. T. (1969). *The hidden dimension*. New York: Anchor books.
- [37] Hargie, O. (2011). *Skilled interpersonal communication. Research, theory and practice*. London: Routledge.
- [38] Hennink, M., Hutter, I., & Bailey, A. (2011). *Qualitative Research Methods*. London: Sage.
- [39] Iacoboni, M. (2009). *Mirroring people. The science of empathy and how we connect with others*. New York: Picador.
- [40] Kelly, S. D., Özyürek, S., & Maris, E. (2009). Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychological Science*, 21 (2) 260–267.
- [41] Kendon, A. (1972). Some relationships between body motion and speech: an analysis of an example. In Siegman, A. Pope, B. (eds.) *Studies in dyadic communication*. New York: Pergamon Press.
- [42] Kendon, A. (1981). Introduction: Current issues in the study of nonverbal communication. In Kendon, A. (ed.) *Nonverbal communication, interaction, and gesture*. The Hague: Mouton Publishers.
- [43] Kendon, A. (2000). Language and gesture: unity or duality? In McNeill, D. (ed.) *Language and gesture*. Cambridge: Cambridge University Press.
- [44] Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- [45] Kita, S. (2000). How representational gestures help speaking. In McNeill, D. (ed.) *Language and gesture*. Cambridge: Cambridge University Press.
- [46] Knapp, M. L., & Hall, J. A. (2006). *Nonverbal communication in human interaction*. Belmont, CA: Thomason Wadsworth.
- [47] Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science* 7, 54–59.
- [48] Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: a process model. In McNeill, D. (ed.) *Language and gesture*. Cambridge: Cambridge University Press.
- [49] Lustig, M. W., & Koester, J. (1993). *Intercultural competence. Interpersonal communication across cultures*. New York: HarperCollins College Publishers.
- [50] McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748.
- [51] McNeill, D. (1992). *Hand and mind: what gestures reveal about thought*. Chicago: University of Chicago Press.

- [52] McNeill, D. (2007). *How language began. Gesture and speech in human evolution*. Cambridge: Cambridge University Press.
- [53] McNeill, D., & Duncan, S. D. (2000). Growth points in thinking-for-speaking. In McNeill, D. (ed.) *Language and gesture*. Cambridge: Cambridge University Press.
- [54] Martin, J. N., & Nakayama, T. K. (2010). *Intercultural communication in contexts*. New York: McGraw-Hill.
- [55] Matsumoto, D., & Hwang, H. S. (2013a). Cultural similarities and differences in emblematic gestures. *Journal of Nonverbal Behavior*, 37, 1–27.
- [56] Matsumoto, D., & Hwang, H. S. (2013b). Cultural influences on nonverbal behavior. In D. Matsumoto, M. G. Frank & H. S. Hwang (eds.), *Nonverbal communication. Science and applications*. London: Sage.
- [57] Matsumoto, D., & Hwang, H. S. (2013c). Body and gesture. In D. Matsumoto, M. G. Frank & H. S. Hwang (eds.), *Nonverbal communication. Science and applications*. London: Sage.
- [58] Poggi, I., & Pelachaud, C. (2008). Persuasion and expressivity of gestures in humans and machines. In Wachsmuth, I., Lenzen, M. & Knoblich, G. (eds.) *Embodied communication in humans and machines*. Oxford: Oxford University Press.
- [59] Proulx, M. J. (2007). *The strategic control of attention in visual search. Top down and bottom-up processes*. Saarbrücken: VDM verlag Dr. Müller.
- [60] Quek, F. et al. (2002). Multimodal human discourse: gesture and speech. *ACM Transactions on Computer-Human Interaction*, 9 (3), 171–193.
- [61] Sobel, C. P. (2001). *The cognitive sciences. An interdisciplinary approach*. Mountain View, CA: Mayfield Publishing Company.
- [62] Sowa, T., Kopp, S., Duncan, S., McNeill, D., & Wachsmuth, I. (2008). Implementing a non-modular theory of language production in an embodied conversational agent. In Wachsmuth, I., Lenzen, M. & Knoblich, G. (eds.) *Embodied communication in humans and machines*. Oxford: Oxford University Press.
- [63] Ting-Toomey, S., & Chung, L. C. (2005). *Understanding intercultural communication*. Los Angeles: Roxbury Publishing Company.
- [64] Tomasello, M. (2008). *Origins of Human communication*. London: A Bradford Book.
- [65] Van Meurs, N., & Spencer-Oatey, H. (2010). Multidisciplinary perspectives on intercultural conflict: the “Bermuda triangle” of conflict, culture and communication. In Matsumoto, D. (ed.) *APA handbook of intercultural communication*. Washington, DC: American Psychological Association.

- [66] Young, R. O. (2011). *How audiences decide. A cognitive approach to business communication*. London: Routledge.