CONCLUSION

Results show that the artificial ITD, which occurs when applying binaural tone Vocoding, affects speech perception. Furthermore, a better SNR is needed for the condition where the artificial ITD is pointing away from the target talker, compared to the condition where it points towards the target talker, for equal speech intelligibility.

Even though young normal-hearing subjects are more sensitive to monaural TFS than elderly hearing-impaired subjects [Hopkins *et al.* 2008], subjects with different age and different hearing status are equally affected by the artificial ITD.

Finally, using binaural tone-vocoding for measuring the benefit of binaural TFS produces side-effects that need to be considered.

REFERENCES

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001**). "The CIPIC HRTF Database" IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pages 99-102.

Andersen, M. R., Kristensen, M. S., Neher, T. and Lunner T. (**2010**). "Effect of Binaural Tone Vocoding on Recognising Target Speech Presented Against Spatially Separated Speech Maskers" Poster at IHCON 2010.

Behrens, T., Neher, T., and Johannesson, R.B. (**2007**). "ERH-42-08-05 Evaluation of a Danish speech corpus for assessment of spatial unmasking" Poster at ISAAR.

Hopkins, K., Moore, B. C. J., and Stone, M. A. (**2008**). "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech" J. Acoust. Soc. Am. **123** (2), 1140-1153.

ISO 7029:**2000**, "Acoustics – Statistical distribution of hearing thresholds as a function of age".

Kuhn, G.F. (**1977**). "Model for the interaural time differences in the azimuthal plane." J. Acoust. Soc. Am. **62** (1), 157-167.

Lunner, T., Hietkamp, R. K., Andersen, M. R., Hopkins, K., and Moore, B. C. J. (**in press**) "Effect of speech material on the benefit of temporal fine structure information in speech for normal-hearing and hearing-impaired subjects" Ear and Hearing.

Moore, B. C. J. and Glasberg, B. R. (**1998**). "Use of a loudness model for hearing aid fitting. I. Linear hearing aids" Br. J. Audioligy. **32**, 317-335.

Speech-inherent functional onomatopoeia as a basis for emotional analysis of phones

JENS BLAUERT

Institute of Communication Acoustics, Ruhr-Universität Bochum, D–44780 Bochum, Germany

Speech sounds (phones) originate in the context of a biological process where the articulators shape the vocal tract into a cascade of cavities and constrictions. This shaping requires muscle activity and these go along with feelings − which the speakers perceive as sitting inside their speech organs (phonetic feelings). The actual feelings depend on the specific form elements that are shaped. Formation of large open cavities associates with a feeling of emptiness, narrow, partly closed cavities are accompanied by the feeling of being pressed, constrictions feel stressed, short thick partitions depressed, long stretched-out partitions filled, satisfied. Each speech sounds designates (symbolically) a specific feeling that is potentially present in the talkers' perception while producing it. In other words, phones represent onomatopoeic acoustic descriptions of the talkers' phonetic feelings. It is considered whether this effect may be exploited, for instance, for word recognition, speaker-emotion recognition, sound design, speech synthesis and sound-quality assessment.

INTRODUCTION

The German philosopher and neurologist *Hans Lungwitz* (1881–1967) has, in 1933, published a monography on the *Psychobiology of Speech/Language* (Psychobiologie der Sprache) as part III of an eight-volume textbook of psychobiology (*Lehrbuch der Psychobiologie*). Due to the confusions of the second world war – *Lungwitz* was suspiciously observed by the Nazies and could hardly present his ideas in public – and since *Lungwitz*'s works are only available in (old fashioned) German, his *Psychobiology of Speech*/*Language* never really caught the attention of his peers. Recently, one of *Lungwitz*'s former students, *Reinhold Becker*, has newly edited the monography, added parts from other volumes of the textbook for better comprehensibility and "refurbished" the language for better readability (Lungwitz 1933, revised 2010). We take this opportunity to report on some fundamental ideas presented in the book, which – although almost eight decades old – may actually still have some relevance, for instance, for modern speech-and-language technology and for sound design. Further, we shall shortly discuss these ideas in the light of the current state of science and technology in the field.

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.

CONCEPTUAL BACKGROUND

Hans Lungwitz is basically a perceptionist, that is, in this respect he agrees with the postulate of *George Berkeley* (Berkeley 1710) that *"esse est percipi"* (to be is to be perceived). However, while *Berkeley* (and other early philosophers) came to this conclusion purely via philosophical reasoning, *Lungwitz* looked for a justification on a biological basis, in other words, on the grounds of natural science. Actually, being a neurologist, he included the brain as the organ of consciousness into his theory and recognized that everything that exists (namely, as perceived) must be "thought" by a brain (Lungwitz 1925). Consequently, he named his novel approach *Psychobiology.* Related lines of thinking can recently be observed, for instance, in (radical) constructivism (see, e.g., Maturana 1987, Dominicus 2011).

FEELINGS

The psychobiological approach is useful for understanding and describing the nature of *"feelings",* a subject for which quite some fuzziness and disagreement can be observed in the relevant literature (compare, e.g., Plutchik 1962, Izard 1977, Campos and Barret 1984, Mees 1985, Frijda 1986, Juslin and Västfjäll 2008, Bergman *et al.* 2009)[1]. From the psychobiological point of view, feelings are one of the categories of percepts of which our world is composed of – besides sensory events (visual, auditory, tactile, etc. "things") and concepts (ideas, thoughts). Feelings exist at their specific time at their specific positions (more or less sharply or diffusely located) with their specific perceptual properties. Their position is, in general, within the borders of the body, but exceptions exist – such as so-called phantom pain after amputation of limbs.

Psychobiology assumes that feelings are associated to specific physiological processes in the body (Lungwitz 1925, Panksepp 1982). After *Lungwitz*, these processes are contractions of muscles in the skeletal, cardial or smooth muscular apparatus. Feelings, when appearing, usually exist right at or close to the position of the associated muscular contraction. Feelings are specific with regard to the type of muscular contraction that they show up together with. *Lungwitz* proposes a system of five elementary feelings, *hunger, fear, pain, sadness* and *gladness*, which go with five specific muscle-contraction patterns as depicted in Fig. 1, taken the muscles around a vessel as an example.[2], namely, contraction of wide round fibres which create an empty space, narrower round fibres which create a narrower empty space, oblique fibres which create a twist, short linear fibres which create a partial release, and

[1] We do not use the term *emotion* in this paper equivocally with feeling, as is often done in literature. See the definitions given in Bergman *et al.* (2009) for differences in the meaning of the two terms.

[2] In the literature various different systems of basic feelings have been proposed with ample congruence but also substantial differences (e.g., Izard 1977, Ekman, *et al.* 1982, Machleidt *et al.* 1989, Ortony & Clore 1989 – see Ortony & Turner 1990 for a synopsis. The latter authors also discuss the question of whether the concept of basic feelings is justified at all.) These disagreements are of no relevance for the body of the current paper, since we use *Lungwitz*'s system only as an ordering system. Other systems would do as well as long as they provide definite links to muscular processes.

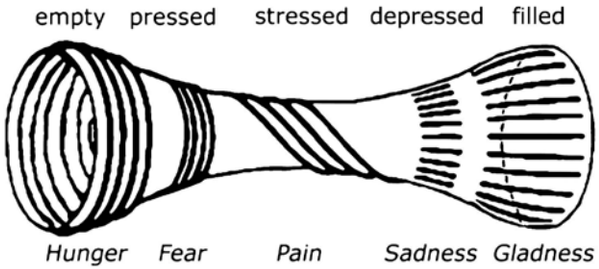

Fig. 1: Schematic of different classes of muscles as covering a vessel or intestine. Upper line: specific physiological states when the muscles have contracted. Lower line: associated feelings (drawing provided by *W. Zabka*)

long linear fibres which create a stretched-out space. In the upper panel of the figure, these physiological states are described in a different wording, referring to the flow of matter in the tube-like structure when these contractions happen, for instance, in the course of peristaltic processes. The lower panel provides terms that describe the feelings as associated with these physiological states. Table I lists further terms to designate these feelings. These are common words in the English language that we use, among others, when we talk to other people about our feelings.

Hunger	**Fear**	**Pain**	**Sadness**	**Gladness**
appetite, courage, desire, emptiness, intention, lust, need, thurst, will, wish, yearning	anxious, ashamed, astounded, careful, cautious, concerned, frightened, harassed, inhibited, insecure, pressed, prudent, timid	beating, cutting, deciding, dividing, drilling, fighting, overcoming, partitioning, peaking, pulsating, impinging, twisting, winding	defeated, depressed, destroyed, disappointed, exhausted, frustrated, grieved, let off, outlawed, regretting, separated, smallness, sorrow, being part only	complete, delighted, free, fulfilled, glad, goal is reached, happy, joyful, mighty, powerful, satisfied, saved, successful, victorious

Table 1: Elementary feelings and various related verbal descriptors

Feelings are valenced, that is, there are feelings that are rated positive and other which are rated negative. Hunger, fear and certain shades of pain are negative in this sense, because they designate the status of the hull and not of its content (emptiness, narrowness, stress), while shades of pain, as well as sadness and gladness are positive, since they designate the matter inside the hull (passed the threshold, arrives piecewise in the new environment, is completed). The valence of the feeling associated to a sound is relevant for sound-quality judgments (Blauert and Jekosch 2010). What holds for the muscular system at large is valid for the vocal tract and the articulators as well. While phone production is happening, feelings are apparent within the speech organ at their specific positions – that is, there exist *phonetic feelings.*

PRODUCTION OF PHONES

To begin with, the following terminological convention shall hold for this article: The words "phone" or, equivocally, "speech sound" are used to designate speech-specific auditory event, that is, percepts that exist as *being heard* and are recognized as elements of speech. In contrast, speech-related acoustic events, that is, mechanical vibrations and waves in/of the speech organ, are denominate by the term "speech wave". The existence of speech waves becomes evident from visual perception, for instance, by looking at a talker's mouth, or tactilely, for example, by putting a finger from outside on his/her larynx and, of course, by adequate instrumental methods. Perception of speech is, consequently, a multi-sensory process.

The process of speech production can be summarized as follows: From the language areas of the cortex, biological signals travel to the speech organ, that is, pneumonic system, larynx, vocal tract and articulators (glottis, velum, mouth/dental opening, tip and/or peak of the tongue). These signals describe in biologically-coded form the concepts stored in the language areas of the cortex that are bound to be expressed orally. These signals act as commands to the speech organ and induce motoric gestures of it. In the course of these gestures, speech waves are generated within the cavities of the vocal tract, whereby excitation can happen periodically at the glottis (voiced excitation) or by turbulences at or sudden opening of narrows of the vocal tract (unvoiced excitation). The speech waves are then radiated to the outside mainly from the oral and nasal openings. In the course of the gestures of speaking, the area function of vocal tract and the position of the articulators are subject to continuous changes as brought about accordingly by series of specific muscle contractions. As mentioned above, the talker perceives this as feelings positioned in the speech organ.

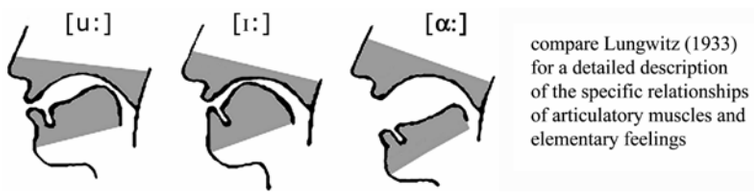

Fig. 2: Vocal-tract profiles for three vowels: [u:] as in "boot", [ɪ:] as in "eve", and [α:] as in "father" (adapted from Flanagan 1972)

A listener can visually observe a talker's facial movements and, eventually, sense the speech production process tactilely. However, in addition, given that the listener is awake and has sound hearing, auditory events will appear in the listener's world, namely a series of phones. Depending on the type of the excitation of the speech organ, the phones will be (voiced) vowels or (voiced or unvoiced) consonants. Phones are usually spatially localized at or close to the talker's mouth opening. Each phone is characteristic of a specific kind of excitation and vocal-tract shape. Figure 2 shows, exemplarily, three vocal-tract profiles for vowels.

PHONES SIGNIFY FEELINGS

In terms of sensory psychology, phones are g*estalts*, that is, sensory entities outlined against other entities, standing in specific relations to these and being characterized by distinct properties. *Hans Lungwitz*, in his book, guides the reader's attention to the fact than one of the particularities of phones is that they have the property of being associated to particular phonetic feelings. He calls this property *Gefühligkeit* (feeling content − literally, *feelingness*). With reference to the elementary feelings listed in Table 1, *Lungwitz* would say that there exist "hungryish", "fearish", "painish", "sadish" and "gladish" speech sounds, among others. The actual type of feelingness of phones is given by the specific muscle contraction and according phonetic feelings during the speech-production process. Following this line of thinking, the following associations of phones to feelings may be stated (Table 2):

	Hunger	**Fear**	**Pain**	**Sadness**	**Gladness**
Vowels	[u:]	[o:]	[ɪ:], [e:]	[uoα], [ə:]	[α:]
Consonants	[h], [m]	[χ], [n]	[r], [j], [s]	[d], [g], [b]	[t], [k], [p]

Table 2: Phones and their association to elementary feelings (a selection)

The important issue about the property of feelingness of auditory events is that listeners can detect it. They actually develop concepts (ideas) regarding the feelings involved in the production process, and in some cases they may even have actual feelings themselves (affection). Recognition of gestures of the speech organ is hypothesized to be an important component of speech recognition (Liberman *et al.* 1967, Galantucci *et al.* 2006) and these gestures may well be better recognized based on the series of feelingness in spoken words and/or phrases.

Further, according to *Lungwitz,* the meaning of many spoken words can be inferred from the specific series of feelings represented by the phones that the words are composed of. For example, the Latin word *MURUS* (wall) denotes the cavity [mu:], the boundary [r] and what is passing in and out, and the suffix [us] indicates the wide environment. As another example, the Greek word *μόνος* (monos – alone, single, segregated) marks what is residing in a cavity–opening–threshold configuration, thus being separated from the environment, where */μον/* denotes this configuration including the fence that separates, and the suffix */ος/* marks what is separated as of male gender. There are many more of such examples in *Lungwitz*'s book and it is hard for a non-phonologist like this author to judge on the justification and precision of the detailed reasoning given in each individual case. Yet, what is relevant in the context of this article is the following: Phones provide information on feelings, strings of phones information on strings of feelings – which helps assign meanings.

In this sense, the gestures of the speech organ describe meanings to be expressed and the series of phones appearing in this context constitute an onomatopoeic representation of strings of phonetic feelings in the talker's world and, consequently, in the listeners' worlds. As far as listeners assign meanings to these strings of phonetic feelings, the string of phones can be conceived as *functional onomatopoeia* of

meanings to be transmitted from the talker to the listeners. By the way, the rôle of prosody – such as pitch contours – is not discussed by *Lungwitz* in this context.

APPLICATION OPPORTUNITIES

From an application point-of-view the most relevant component of *Lungwitz*'s theory is the statement that feelings are specifically associated with physiological states of the muscular system. This would mean that when we know the physiological states, we can estimate the feeling with high precision and vice-versa, namely,

- Phonetic feelings can be predicted from the area function of the vocal tract and the positions of the articulators
- The area function of the vocal tract and the position of the articulators can be estimated from the phonetic feelings

Listeners perceive the phonetic feeling via the phones that they hear, since these phones carry information about the feeling due to their property of feelingness.

- Phonetic feelings can be recognized by listeners via the phones that they hear

Strings of phonetic feeling are descriptors of motoric gestures of the speech organ, the identification of which supports word recognition. This may also be helpful for the assignment of meaning to words.

- Word recognition and assignment of meaning to speech is supported by the identification of phonetic feelings

Feelings are valenced as positive or negative. Sound quality is co-determined by the valence of the feelings associated to the sounds to be judged upon.

- Knowledge of the feelings associated to the physiological states of the speech organs allows for estimations of speech and voice quality
- Knowledge of the valence of feelings is helpful for the design of phones – and sounds in general – with specified qualities

Lungwitz's theory, if confirmed, would open application opportunities, for instance, in automatic recognition of speaker emotion, speech-sound design, instrumental speech-synthesis, meaning assignment to phones and words, and assessment of speech-sound quality in speech-recognition and speech-comprehension technology.

DISCUSSION

Lungwitz's theory has been buried in the archives for more than 75 years but seems to be worthy of being reintroduced into scientific discussion for further evaluation. Its most prominent deficiency is a lack of experimental data, since *Lungwitz* himself – in the tradition of early German psychology – has developed his theories mainly on the grounds of personal observation (introspection). Machleidt *et al.* (1989) tried to provide supportive clues with EEG analysis, but hard experimental evidence of the existence of *Lungwitz*'s basic-feelings system is still not available.

That phones transmit information about feelings is not a new idea and widely accepted. However that phonetic feelings are represented by them, seems to be a novel observation. Phones, to be sure, are commonly understood as sign carriers in the information sciences. Semiotics analyses this fact in more detail and states that signs

can be either indices or icons or symbols (Jekosch 2005, Blauert and Jekosch 2010). In the context of this paper, their function as index is the most relevant one, since the phones signify motoric states of the speech organ and thus support the identification of gestures of this organ. The *motor theory* of speech perception (Liberman *et al.* 1967, Liberman and Mattingly 1989, Galantucci *et al.* 2006) confirms the relevance of these processes for speech recognition.

Starting from the indexical function of phones, we can retrace *Lungwitz*'s hypothesis as to which the string of phones of a word – and thus, the series of motoric actions of the speech organ – are biologically related to the meaning that the cortex of the talker is bound to express. Similar ideas have been considered in the context of the phylogeny of language (e.g., Lenneberg 1967), but there seems to be wide agreement among experts that they cannot be extrapolated to word meanings in general.

The postulate of *Lungwitz* as to which listeners can detect the feelingness of phones and, consequently, develop their own conceptual image of these feelings – or even have actual feeling of similar kind as those of the talker – is supported by the recent discovery of *mirror neurons* in the nervous system (e.g., Di Pellegrino *et al.*, 1991, Rizzolati and Craighero 2004), but needs further experimental verification as well.

REFERENCES

Bergman, P., Sköld, P., Västfjäll, D., and Fransson, N. (**2009**). "Perceptual and emotional categorization of sound" J. Acoust. Soc. Am. **126**, 3156–3167.

Berkeley, G. (**1710**). *Treatise on the principles of human knowledge*. Jeremy Pepyat, EIR−Dublin.

Blauert, J., and Jekosch, U. (**2010**). "A layer model of sound quality" In: Proc. 3rd Int. Conf. Perceptual Quality of Systems, D−Bautzen (to appear also in J. Audio-Eng. Soc. **2011**).

Campos, J. J., and Barret, K. C. (**1984**). "Towards a new understanding of emotions and their development" In: Izard, C. E., Kagan, J., and Zajonc, R. B. (eds.):, *Emotion, cognition, and behavior*, 229–236, Cambridge Univ. Press, New York.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (**1992**). **"**Understanding motor events: a neurophysiological study" Exper. Brain Res. **91**, 176–80.

Dominicus, R.-D. (**2010**). *Radikaler Konstruktivismus versus Realismus.* Diplomica Verlag, D–Hamburg.

Ekman, P., Friesen, W. V., and Ellworth, P. (**1982**). "What emotion categories or dimensions can be observed from facial behaviour?" In: Ekman, P. (ed.) *Emotion in the human face*, pp. 39–55. Cambridge Univ. Press, New York NY.

Flanagan, J.L. (**1972**). *Speech Analysis, Synthesis and Perception*, 2nd ed., Springer**,** Berlin−Heidelberg−New York NY.

Frijda, N.H. (**1986**). *The emotions.* Cambridge Univ. Press, NewYork NY.

Galantucci, B. Fowler, C. A., and Turvey, M-T. (**2006**). "The motor theory of speech perception reviewed" Psychon. Bull. Rev. **13**, 361–377

Izard, C. C. E. (**1977**). *Human emotions.* Plenum Press, New York NY.

Jekosch, U. (**2005**). "Assigning meaning to sounds – semiotics in the context of product-sound design" in: J. Blauert (ed.): *Communication Acoustics*, pp. 193–219, Springer, Berlin–Heidelberg–New York NY.

Juslin, N. P. and Västfjäll, D. (**2008**). "Emotional responses to music: The need to consider underlying mechanisms" Behav. and Brain Sciences **31**, 559–621

Lenneberg, H. E. (**1967**). *Biologische Grundlagen der Sprache* (biological foundations of language). Suhrkamp, D-Frankfurt/Main.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert–Kennedy, M. (**1967**). "Perception of speech code". Psychol. Rev. **74**, 431-461.

Libermann, A. M., and Mattingly, I. G. **(1989)** "A specialization for speech perception" Science **243**, 489–494.

Lungwitz, H. (**1925**). *Die Entdeckung der Seele – Allgemeine Psychobiologie* (the discovery of the psyche − general psychobiology). De Gruyter, D–Berlin (cited here: 5th edition, De Gruyter, D–Berlin, **1947**).

Lungwitz, H. (**1933**). *Die Psychobiologie der Sprache* (the psychobiology of speech/language). Brückeverlag Kurt Schmersow, D–Kirchhain (cited here: 3rd, revised edition, Becker, R., ed., Thieme, Stuttgart–New York NY, **2010**).

Machleidt, W., Gutjahr, L., and Mügge, A: (**1989**). *Grundgefühle* (basic feelings). Springer, Berlin–Heidelberg–New York NY.

Maturana, H. R.: (**1987**). "Biology of language: the epistemology of reality" In: Miller, G.A., and Lenneberg. E. (eds.), *Psychology and biology of language and thought,* pp. 27−63. Academic Press, New York NY.

Mees, U. (**1985**). "What do we mean when we speak of feelings? On the psychological texture of words denoting emotions" Sprache und Kognition **1**, 2–20.

Ortony, A. and Clore, G. L. (**1989**). "Emotions, moods and conscious awareness". Cogn. & Emotion **3**, 125–137.

Ortony, A. and Turner, T. J. (**1990**). "What's basic about basic emotions" Psych. Rev. **97**, 315–331.

Panksepp, J. (**1982**). "Towards a general psychobiological theory of emotions". Behav. and brains sciences. **5**, 407–467.

Plutchik, R. (**1962**). *The emotions: facts, theories, and a new model.* Random House, New York NY.

Rizzolatti, G. and Craighero, L. (**2004**). "The mirror-neuron system". Ann. Rev. Neuroscience **27**, 169–192

Audiovisual integration in speech perception: a multi-stage process

KASPER ESKELUND[1] , JYRKI TUOMAINEN[2] AND TOBIAS ANDERSEN[1]

[1] *Cognitive Systems, Department of Informatics and Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark*

[2] *Speech Hearing and Language Sciences, University College London, UK*

Integration of speech signals from ear and eye is a well-known feature of speech perception. This is evidenced by the McGurk illusion in which visual speech alters auditory speech perception and by the advantage observed in auditory speech detection when a visual signal is present. Here we investigate whether the integration of auditory and visual speech observed in these two audiovisual integration effects are specific traits of speech perception. We further ask whether audiovisual integration is undertaken in a single processing stage or multiple processing stages.

INTRODUCTION

Integration effects such as the McGurk effect (McGurk and MacDonald, 1976) and the detection advantage associated with audiovisual speech (Grant and Seitz, 2000) show that vision and hearing are integrated in speech perception. It is, however, unknown whether the processes underlying such audiovisual integration are specific for perception of speech, or if they pertain to audiovisual perception in general. Moreover, audiovisual integration is often tacitly assumed to be undertaken in a single step (Massaro, 1998; Vatakis and Spence, 2007). In the experiment reported here, we test whether audiovisual integration as seen in the McGurk effect and the audiovisual detection advantage occurs for both non-speech and speech perception. We further test these integration effects as to investigate whether they show different properties in non-speech and speech conditions. If the latter is the case, it may indicate that the effects are related to dissociated processes supporting the claim that audiovisual integration of speech is multi-faceted.

Grant and Seitz (2000) showed that seeing a synchronous visual speech signal is advantageous when detecting an acoustic speech signal masked by noise. Presenting three sentences in audiovisual and auditory-only formats masked by acoustic noise, they found that the advantage associated with the presence of the visual speech signal in the audiovisual stimulus was equivalent to a 1.6 dB gain of the auditory-only stimulus. Investigating the dynamics of the acoustic and visual stimuli, they showed that the magnitude of the advantage depends on the degree of correlation between changes in lip opening area and sound intensity. On this basis, they proposed the *peak listening hypothesis*, stating that cues in the visual signal guides

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.