# ROBUSTNESS AS A DESIGN FEATURE OF SPEECH COMMUNICATION

BODO WINTER

*Dept. of Cognitive and Information Sciences, University of California, Merced,*
*5200 North Lake Rd., Merced, CA, 95343, USA*

MORTEN H. CHRISTIANSEN

*Department of Psychology, Cornell University, Ithaca, NY, 14853, USA*
*Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM, 87501, USA*

As a communication system, human language is incredibly robust. In this paper, we identify some of the mechanisms that make speech communication robust in the face of noise and variation. We argue that many phenomena that are usually studied for their own sake (speech adaptation, accommodation, distributed redundancy) work together in order to increase the robustness of speech. Robustness can thus be used as a unifying concept for a seemingly disparate collection of experimental data. We outline the implications of robustness for the study of language evolution, and we discuss how both biological and cultural evolution may have played a role in making language robust. Taking the robustness of language into account makes it clear that not only do linguistic systems evolve with respect to functional considerations, but also with respect to how those functions can be maintained in the light of a multitude of perturbations.

## 1. Introduction

Robustness is the property of a system that enables it to cope with noise and interference; it is what allows a system to maintain its function despite internal and external perturbations. The concept has received a lot of attention in complexity science and biology (for a review, see Kitano, 2004), but it has received considerably less attention in language research – despite the fact that language obviously seems to have some degree of robustness, otherwise we wouldn’t be able to talk in a variety of acoustic environments and with varying degrees of environmental noise. Moreover, linguistic communication even works despite a large amount of variation: each and every utterance is different along some dimension. Speakers differ with respect to sex, age, dialect, personality and many other characteristics that affect speech patterns (inter-speaker variation), and utterances differ from one another because of motor variability (intra-speaker variation). The presence of robustness raises key questions for language evolution research: what are the specific properties that make it possible for us to communicate in the face of noise and variation, and how did these evolve?

Looking at speech communication in particular, it is possible to uncover how many different parts of language work together to safeguard the communication process against perturbations. However, rather than there being a hotchpotch of different unconnected properties that increase robustness, we argue that the different properties can be characterized as a small set of **robustness-enhancing features** over which generalizations can be made:

1. Speech adaptation and motor equivalence
2. Speech accommodation
3. Distributed redundancy

This list is not meant to be exhaustive. For now, we focus on the robustness of the speech signal, but note that there are additional robustness enhancers at other levels of linguistic description, e.g., there is considerable redundancy in morphosyntax (bipartite morphemes, participant roles that are encoded simultaneously through word order and case, etc.). Moreover, we focus on the perspective of speech production, as others have already explored the robustness of speech perception (e.g., Diehl, 2011). In the following, we discuss the theoretical and empirical support for each of the three robustness-enhancing features in turn.

## 2. Speech adaptation

In this paper, we use the term “adaptation” not in its evolutionary sense, but to describe the observation that language users are able to rapidly adapt to the onset of environmental noise (e.g., Grynpas et al., 2011). Similarly, speakers are able to react appropriately to temporary or long-term mechanical speech impediments (e.g., Tremblay et al., 2003; Bressman, 2006). Speech adaptation is a feedback mechanism that detects changes in the acoustic output of one’s own speech and quickly alters speech production in order to arrive at a more intelligible output.

Experimental evidence for speech adaptation comes from auditory feedback studies, where speakers that produce speech are played back a synthetically altered version of their own speech (e.g., shift in pitch). With this paradigm, it was shown that people are able to adapt to changes in loudness (e.g., Lane &

Tranel, 1971), fundamental frequency (Jones & Munhall, 2000) and formant frequencies (Tremblay et al., 2003). Speech adaptation acts both short-term in a rapid and automatic fashion, as well as on a long-term basis, and it also works with feedback from the sensorimotor system (Tremblay et al., 2003). The potential power of speech adaptation as a robustness enhancer is shown by the fact that intelligibility is assured even despite such extreme impediments as tongue splits (Bressman, 2006).

Speech adaptation is closely linked to the notion of motor equivalence, the idea that there are multiple ways in which to produce acoustic outputs that have identical linguistic effects on the listener. This idea comes from early experiments where participants were still able to produce intelligible speech despite mechanical interference due to speech adaptation (e.g., Folkins & Abbs, 1975). One perspective on this data is that speech sounds tend to be within so-called "quantal" regions (Stevens, 1979)—that is, regions within the articulatory apparatus that allow a high degree of motor variation while keeping the amount of acoustic variation at a minimum. Having speech patterns within these quantal regions not only assures robustness against external perturbations, but also against the inherent degree of variability in speech movements (gestures will almost always miss a given articulatory target region to a certain degree).

The evolutionary significance of speech adaptation is two-fold. First, given that adaptation only works if there is some degree of motor equivalence, it becomes important to explain how speech sounds came to exhibit this property. We hypothesize that there may be evolutionary pressures in cultural language evolution towards higher degrees of motor equivalence. Speech sounds which are in articulatory regions that do not afford a lot of variation will tend to be more acoustically variable – and this inevitably leads to higher degrees of misperception. Sounds that are often misperceived are unstable in diachronic terms, and will either disappear or develop towards a pattern that is less prone to misperception (Blevins, 2004). Preliminary support for such a process can be gleaned from simulations which show that sound systems develop towards perceptually more distinct regions (e.g., de Boer, 2001), however, we are currently unaware of computational simulations that specifically address the role of motor equivalence.

Second, the fact that people are able to perform adaptation rapidly, automatically and non-consciously suggests that there may also be a biological component to this ability, thus opening the possibility for biological evolution having played a role (e.g., speakers who were better at speech adaption may have had a selective advantage). Alternatively, the speech adaptation mechanism might be a spandrel, e.g., the ability to perform adaptation might be a simple outgrowth from our general capacity to imitate, which may be a prerequisite for language (e.g., Donald, 2005). Intriguingly, speech adaptation seems to have corollaries in other species. House finches, for example, perform rapid real-time frequency shifts of their songs in response to urban noise (Bermúdez-Cuamatzin et al., 2011). The presence of behavioral adaptation mechanisms in other species may allow us to address its possible connection to imitation, as well as whether it is a spandrel or a biological adaptation.

## 3. Speech accommodation

Whereas speech adaptation is a response to changes in one's own speech patterns, accommodation refers to changes with respect to the speech patterns of others. People can either increase or decrease similarity of their speech patterns (convergence and divergence, respectively), however, speech convergence seems to be much more common than divergence (e.g., Giles & Coupland, 1991: 66).

The major theory that seeks to explain accommodation behavior, Communication Accommodation Theory (CAT, e.g., Giles & Coupland, 1991), grounds speech convergence in the common desire to decrease social distance with our interlocutors. While this is very much *one* likely source of speech convergence behavior, other evidence indicates that speech convergence cannot be a solely social phenomenon; e.g., speakers rapidly converge to recordings of nonwords (e.g., Goldinger, 1998), individual syllables presented out of context (Nielsen, 2005), and to computerized agents even if they do not feel sympathetic to those agents (Staum Casasanto et al., 2010). Thus, we suggest, from an evolutionary perspective, that speech convergence also serves to increase robustness. By having a mechanism through which we can increase the similarity of the communicative code over repeated interactions, we can deal with the diversity of different speakers.

Just like speech adaptation, accommodation is a rapid, automatic and often non-conscious process. Hence, the question arises whether speech accommodation simply follows from having the (biological) ability to imitate, or whether there is a specific biological adaptation towards being better at accommodating. Again, there are important and interesting corollaries in other species, which possibly allow testing these views: both birds (e.g., Mammen & Nowicki, 1981) and non-human primates (Lemasson & Hausberger, 2004) exhibit evidence for the convergence of communicative codes.

## 4. Distributed redundancy

Many researchers have argued that language has some degree of redundancy (e.g., Pinker & Jackendoff, 2005). In speech, redundancy is ubiquitous. For example, stress is simultaneously cued by pitch, loudness, duration and spectral slope. Utterance endings are cued by a loss of energy, a reduction in pitch, final lengthening and sometimes a following pause. Word boundaries are cued by (depending on the language) transition probabilities, phonotactics, stress, word-final lengthening, word-final phonological processes and sometimes a following pause. Evidence that this **cue layering** actually increases robustness comes from cut-off studies (e.g., Cho, 1996), where listeners have fewer cues available and then use the other cues, or from noise overlay studies (e.g., Xu et al., 2005).

Many of these cues arise epiphenomenally because of simple biomechanical or physical reasons. For example, the fact that pitch is increased after aspirated as opposed to non-aspirated stops has to do with the higher amount of glottal airflow which then leads to a faster glottal vibration. Due to the increased flow, the glottis cannot help but to vibrate faster after aspirated stops and thus, this additional cue comes “for free”. Whether a correlate in speech production is then actually used as a cue depends on the perceiver, but as Moreton (2004: 13) says, “correlates of a contrast are typically cues to that contrast.”

However, redundancy in and by itself is not enough to facilitate language use; the redundancy also needs to be **distributed** in the temporal domain. Very loud broadband noise would be able to selectively interfere with all phonetic cues at a given point in time, but if the information is spread out over the signal, misperception at a particular point in time can be dealt with by retrieving the information later. This kind of temporal distribution is ubiquitous in phonetic systems, e.g., Cho (1996) showed that Korean listeners are able to decide whether Korean stops are tensed, lax or aspirated purely based on voice quality differences in the following vowel. Xu and colleagues (2005) showed that people are able to locate the position of sentential focus even well after the focused constituent because of cues that follow the constituent (called “post-focus compression”).

The sound systems of the world’s languages seem to exhibit distributed redundancies to a great extent. We take this as an indication that there potentially are selective pressures towards languages being more redundant. These pressures could manifest themselves through cultural language evolution. This view is in line with recent trends in phonology, where different researchers have started to focus on the evolution of sound patterns through perceptual and articulatory constraints. Accordingly, Blevins (2004) argues that the sound patterns that are most frequently observed in the languages of the world are those that are easy to produce and easy to perceive. Part of what makes a phoneme or a contrast easy to perceive is determined by the amount of redundancy and by the spread of this redundancy. The idea that language as a system converges on higher degrees of redundancy leads to the testable predictions for linguistic typology: the most frequent sound patterns of the world should also be the most redundant ones.

## 5. Conclusions

Our list of robustness enhancers is meant to be a first step towards thinking about robustness as a general property of speech communication—a property that likely extends to other areas of language as well. Because the enhancing features that we identified are present (to some degree) in *all* languages, we hypothesize that robustness may be a **design feature** of language and thus of similar importance to the evolution of language as arbitrariness and duality of patterning (cf. Hockett, 1960). Moreover, the concept of robustness provides a unifying perspective on a seemingly diverse set of descriptive and experimental data, ranging from the speech production studies that demonstrate redundancy to experiments on adaptation and accommodation.

If robustness is an important feature of language, the question naturally arises as to what its origin may be. As outlined in the sections above, both cultural evolution and biological evolution may have played a role. Speech adaptation and speech convergence via accommodation seem to derive from the biology of the speakers of a language. However, this leaves unresolved whether there were specific adaptations towards these skills, or whether these capabilities simply follow from our more general capability to imitate and therefore may be a reflection of prior biological abilities. One avenue into investigating the biological nature of these capabilities is by looking at other species (which sometimes show signs of adaptation and accommodation), another is by conducting individual differences studies (see, e.g., Misyak & Christiansen, in press). For example, the view that speech accommodation is a spandrel that follows from our imitation capacity makes the prediction that people who are better at implicitly imitating (also with respect to non-linguistic behaviors) should more readily perform speech convergence when talking to others.

With respect to motor equivalence, functional redundancy and its temporal distribution, we suggest that cultural evolution may be the predominant factor that produces these robustness-enhancing features. From the simple assumption that speakers tend to stick with those patterns that are most easily perceived, and best received in face of perturbations, it follows that speech systems should evolve in the direction of more “quantal” regions of the speech apparatus and towards more distributed redundancy.

We conclude by noting that the evolutionary hypothesis we have forwarded in this paper has primarily relied on reinterpretations of existing results in the speech literature. However, our notion of robustness also allows us to derive new testable hypotheses. For example, convergence of speech patterns between interlocutors should be stronger in noisy environments. Or, if participants are asked to interact with one another using an artificial language, then we would expect that across multiple experimental generations, they would show a tendency towards using forms with higher degrees of redundancy and more temporally distributed cues. Hence, thinking about language from the perspective of robustness and making robustness a topic for further investigation may prove very fruitful for future research into the evolution of language.

## References

Bermúdez-Cuamatzin, E., Ríos-Chelén, A. A., Gil, D., & Garcia, C. M. (2011). Experimental evidence for real-time song frequency shift in response to urban noise in a passerine bird. *Biology Letters*, 7, 36-38.

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Bressmann, T. (2006). Speech adaptation to a self-inflicted cosmetic tongue split: Perceptual and ultrasonographic analysis. *Clinical Linguistics & Phonetics*, 20, 205-210.

Staum Casasanto, L., Jasmin, K., & Casasanto, D. (2010). Virtually accommodating: Speech rate accommodation to a virtual interlocutor. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 127-132). Austin: Cognitive Science Society.

Cho, T. (1996). Vowel correlates to consonant phonation: an acoustic-perceptual study of Korean obstruents. M.A. thesis, University of Texas at Arlington.

Diehl, R. L. (2011). On the robustness of speech perception. *International Congress of Phonetic Science,* Hong Kong.

Donald, M. (2005). Imitation and mimesis. In S. Hurley & N. Chater (Eds.), *Perspectives on Imitation: From Neuroscience to Social Science, Volume 2: Imitation, Human Development, and Culture* (pp. 282-300). Cambridge, MA: MIT Press.

Folkins, J. W., & Abbs, J.H. (1975). Lip and jaw motor control during speech: Responses to resistive loading of the jaw. *Journal of Speech and Hearing Research*, 18, 207-220.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review,* 1052, 251-279.

Giles, H. (1973). Accent mobility: a model and some data. *Anthropological Linguistics*, 15, 87-109.

Giles, H., & Coupland, N. (1991). *Language: Contexts and consequences*. Pacific Grove: Brooks/Cole.

Grynpas, J., Baker, R., & Hazan, V. (2011). Clear speech strategies and speech perception in adverse listening conditions. *International Congress of Phonetic Science*, Hong Kong.

Hockett, C. (1960). The origin of speech. *Scientific American*, 203, 88-111.

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *Journal of the Acoustical Society of America*, 108, 1246–1251.

Kitano, H. (2004). Biological Robustness. *Nature*, 5, 826-837.

Lane, H. L., & Tranel, B. W. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech Language and Hearing Science*, 14, 677-709.

Lemasson, A., & Hausberger, M. (2004). Patterns of vocal sharing and social dynamics in a captive group of Campbell's monkeys (Cercopithecus campbelli campbelli). *Journal of Comparative Psychology*, 118, 347-359.

Mammen, D. L., & Nowicki, S. (1981). Individual differences and within-flock convergence in chickadee calls. *Behavioral Ecology and Sociobiology*, 9, 179-186.

Misyak, J. B., & Christiansen, M. H. (in press). Statistical learning and language: An individual differences study. *Language Learning.*

Moreton, E. (2004). Realization of the English postvocalic [voice] contrast in F1 and F2. *Journal of Phonetics*, 32, 1-33.

Nielsen, K. (2005). Specificity and generalizability of spontaneous phonetic imitation. *Proceedings of the 9th International Conference on Spoken Language Processing,* Pittsburgh, USA.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, 19, 2382-2392.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition*, 95, 201-236.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17, 3-45.

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*, 423, 866-869.

Xu, Y., Xu, C.X., & Sun, X. (2004). On the temporal domain of focus. *Speech Prosody*, Nara, Japan, *81-84.*