Information Density and the Predictability of Phonetic Structure

Bernd Möbius, Bistra Andreeva, Ivan Yuen, Omnia Ibrahim*

Abstract. This work is concerned with the relation between information density and linguistic encoding in phonetics and human speech processing. Information density of a linguistic unit is defined in terms of surprisal (the unit's negative log probability in a given context). The main hypothesis underlying our experimental and modelling work is that speakers modulate details of the phonetic encoding in the service of maintaining a balance of the complementary relation between information density and phonetic encoding. So far we have investigated effects on subword (segmental and syllable) levels and in local prosodic structures (at phrase boundaries).

Keywords. phonetics, information density, surprisal, contextual predictabilty, segments, syllables, prosodic boundaries, speech in noise

1. Introduction. The working hypothesis underlying our research is that speakers modulate details of the phonetic signal in the service of maintaining a balance of the complementary relation between surprisal (or linguistic information density, or contextual (un)predictability; see section 2.1 for a formal definition) and phonetic encoding. To test this hypothesis we analysed the effects of surprisal on phonetic encoding, in particular on dynamic vowel formant trajectories, stop consonant voicing, syllable duration, and vowel space size, while controlling for several basic factors related to the prosodic structure, viz. lexical stress and major prosodic boundaries, in the statistical models that accounted for phonetic effects of changes in surprisal (e.g. Malisz et al. 2018, Brandt et al. 2021). Our findings are generally compatible with a weak version of the Smooth Signal Redundancy (SSR) hypothesis (Aylett & Turk 2004, 2006, Turk 2010), suggesting that the prosodic structure mediates between requirements of efficient communication and the speech signal. However, this mediation is not perfect, as we found evidence for additional, direct effects of changes in predictability on the phonetic structure of utterances. These effects appear to be stable across different speech rates in models fit to data derived from six different European languages (Malisz et al. 2018).

Furthermore, we investigated the interaction of the prosodic hierarchy with surprisal, rather than including prosodic factors merely as control variables in the statistical analysis. This turned out to be a fruitful paradigm, especially with respect to predictions based on the SSR hypothesis. As expected, an increased pre-boundary syllable duration in high-surprisal contexts was observed above and beyond the well-established phrase-final lengthening effect (Andreeva et al. 2020). Moreover, we investigated possible interactions between speech enhancements arising from surprisal with those occurring in adverse acoustic conditions, i.e., in the presence of a noisy channel. Again, speakers behave strategically by making an effort to increase the difference between syllables in high vs. low surprisal contexts in the presence of noise (Ibrahim et al. 2022b). Finally, we explored the degree to which non-native speakers consider surprisal patterns of the target language in their L2 productions, finding that advanced but not intermediate L2 speakers take target-like surprisal patterns into account (Brandt et al. 2019).

*All authors: Language Science and Technology, Saarland University. Email: moebius@lst.uni-saarland.de & andreeva@lst.uni-saarland.de & ivyuen@lst.uni-saarland.de & omnia@lst.uni-saarland.de.

2. Methodology.

2.1. LANGUAGE MODELING. In all studies reported below, we estimated syllable- and segment-based surprisal from language models based on the deWaC (deutsche Web as Corpus kool yinitiative) corpus (Baroni et al. 2009). The corpus is a collection of web-crawled data containing about 1.7 billion word tokens and 8 million word types from a diverse range of genres such as newspaper articles and chat messages. The corpus was pre-processed and normalized using German Festival (Möhler et al. 2000). This procedure consisted of removing unnecessary, irrelevant or duplicate document information, for example, web-specific structures such as HTML structures or long lists etc. After pre-processing, the normalized corpus was divided into a training set (80%) and a test set (20%). Syllable-based trigram language models including word boundary as a unit and segment-based trigram models including syllable boundary as a unit were trained on the training set using the SRILM toolkit (Stolcke 2002). All language models underwent Witten-Bell smoothing (Witten & Bell 1991). The best-performing trained language model was then used to calculate the conditional probability of a syllable or segment, given the preceding context, which is defined as the two preceding units including word or syllable boundary (i.e., $S(unit_i) = -log_2 P(unit_i|unit_{i-1}, unit_{i-2})$, where S = surprisal and P = probability (Hale 2016). The conditional probabilities thus constituted the surprisal measures for the target syllables and segments.

2.2. SPEECH DATA SETS. With one exception, all studies reported in this paper relied on found corpora comprising speech data recorded for entirely different purposes by other research groups, in some cases providing automatically or manually produced linguistic annotations.

The speech data set underlying the formant trajectory analysis (section 3.1.1) is the Siemens Synthesis corpus (SI1000P) (Schiel 1997). These recordings were done to provide high-quality material for concatenative speech synthesis. The corpus contains audio recordings from two professional, middle-aged, male speakers of Standard German. They were asked to read as if in a broadcasting setting. The speakers recorded the same 992 sentences selected from the Frankfurter Allgemeine newspaper corpus (SI1000) in a sound-treated studio at a sampling rate of 48 kHz and 16 bits, filtered and down-sampled to 16 kHz. Canonical transcriptions and automatic word and phoneme segmentations are available.

For the studies on stop consonant voicing (section 3.1.2) and on the effect of channel characteristics on speech units with different degrees of surprisal (section 3.2), we created a Lombard corpus of German read speech. A set of 60 sentences was selected from a version of the deWaC corpus (Baroni et al. 2009) based on high vs. low surprisal bins. The selected sentences were shortened so that they could be displayed in two lines on the screen while preserving the target syllable's surprisal value. Each target CV syllable was part of a polysyllabic word in a sentence, where C beginning with a /p, b, d, k/ plosive was combined with 5 vowels (/a:, e:, i:, o:, u:/). The order of the noise conditions (0 dB vs. -10 dB SNR) was counterbalanced, retaining the middle block for the no-noise condition (baseline). This design was chosen to reduce the tendency of speakers to linearly adjust to noise. Productions were recorded and stored as a mono .wav file with a sampling rate of 48 kHz and 24 bits per sample.

Finally, for the study on the—possibly interacting—effects of surprisal and prosodic structure on phrase-final syllables (section 3.3) we used a subset of the German Discourse Information Radio

News Database for Linguistic analysis (DIRNDL) (Eckart et al. 2012, Börkelund et al. 2014). The corpus consists of approximately five hours of read speech produced by nine news anchors (5m, 4f). It was automatically segmented into words, syllables, and phonemes. Pitch accents and prosodic boundaries were manually labeled by one student assistant according to GToBI(S) (Mayer 1995). Word level annotations were mapped to syllable-based prosodic labels using Festival (Möhler et al. 2000).

The exception to this pattern of analyzing found speech data sets was the study of information density as a possible measure of language proficiency (section 3.4), for which the participants (six Bulgarian L2 speakers at B2 proficiency level, six Bulgarian L2 speakers at C1 proficiency level, and six German native speakers) read aloud German text passages from the EUROM-1 corpus (Chan et al. 1995).

3. Analysis of surprisal in subword units.

3.1. Effects of surprisal in the production and perception of segments.

3.1.1. Formant dynamics. Spectral expansion due to predictability effects can not only be observed in single, pointwise measures of the spectrum but also in formant trajectories. Easily predictable German vowels have less formant change, less F1 slope and velocity, and are less curved in their F2 than vowels that are less predictable. Vowels in low-surprisal (high-frequency) words show less vowel-inherent spectral change. Conversely, high-surprisal units are produced with an expanded range of values on several phonetic features (Brandt et al. 2019, 2021).

We investigated whether variability in German formant trajectories can be explained by surprisal and prominence (i.e., primary lexical stress), as well as an interaction of both factors. We included word frequency as an additional information-theoretic measure in our models. We used generalized additive mixed models (GAMMs) to compare the shape of formant trajectories in different surprisal contexts. GAMMs combine parametric terms and smooth terms in their structure, that is, they allow investigation of the relations between a response and one or more covariates in average values and also in nonlinear terms. In addition, they incorporate random effects, that is, random intercepts, slopes, and smooths. Surprisal values were based on the biphone or triphone of the preceding or following context of the vowel. Only monophthongs in content words were considered in the study.

For average F1 and F2, we found expected results for different vowel phonemes within the acoustic vowel space. The significant interaction effect between the factors vowel and stress in the F1 and F2 models confirmed that vowels in the stressed position are more dispersed in the vowel space than vowels in the unstressed position. We were particularly interested in the results of the smooth terms including surprisal. The GAMM output showed that the first and second formant trajectories in German are affected by surprisal based on both the preceding and the following context, and by the interaction of surprisal and stress. Unstressed vowels seem to show higher variability in their formant trajectory at different surprisal levels than stressed vowels. Differences in average formant values are also more readily expressed as a function of surprisal in stressed vowels than in unstressed vowels. Our results show that effects of contextual predictability on formant variability are not limited to pointwise measurements of the vowel, as seen in studies on the effect of predictability on vowel dispersion (Malisz et al. 2018), but affect the dynamics throughout the entire vowel duration. When interpreted against the background of the uniform information density hypothesis (Levy & Jaeger 2006), our findings add to the concept that the rational speaker uses optimization strategies in speech production throughout the entire utterance to ensure successful communication. This strategic behaviour of the speaker also has an effect on the characteristics of formant movement and is observed while controlling for linguistic factors that are known to affect formant movement, such as vowel duration or phonetic context (Brandt et al. 2021).

3.1.2. VOICING OF STOP CONSONANTS. Cross-linguistic evidence suggests that syllables in predictable contexts have shorter duration than in unpredictable contexts. However, it is not clear whether predictability uniformly affects phonetic cues of a phonological feature in a segment. In this study (Ibrahim et al. 2022a) we explored the effect of syllable-based predictability on the durational correlates of the phonological stop voicing contrast in German, viz. voice onset time (VOT) and closure duration (CD). The target stop consonants /b, p, d, k/ occurred in stressed CV syllables in polysyllabic words embedded in a sentence, with either voiced or voiceless preceding contexts. The syllable occurred in either a low or a high predictable condition, which was based on a syllable-level trigram language model. We measured VOT and CD of the target consonants (voiced vs. voiceless). The results of our statistical analysis using linear mixed effects models showed an interaction effect of predictability and the voicing status of the target consonants on VOT, but a uniform effect on closure duration. This interaction effect on a primary cue like VOT indicates a selective effect of predictability on VOT, but not on CD. This suggests that the effect of predictability is sensitive to the phonological relevance of a language-specific phonetic cue.

3.2. LOMBARD SPEECH MODIFICATIONS INTERACTING WITH SURPRISAL.

Speakers tend to speak clearly in noisy environments, while they tend to reserve effort by shortening word duration in predictable contexts. It is unclear how these two communicative demands are met. Despite the variety of studies on speech enhancement strategies in noisy conditions (for a review, see Cooke et al. 2014), these studies have focused on the Lombard effect or the predictability effect separately while our work extended previous literature by examining both factors in tandem. Our study investigated whether surprisal or contextual unpredictability is modulated by different levels of background white noise.

We analyzed the effects of noisy channel characteristics on speakers' productions, in particular the acoustic realizations of syllables in predictable vs. unpredictable contexts across different background noise levels (Ibrahim et al. 2022b). A set of 60 CV syllables divided into two surprisal groups (high vs. low) and crossed with three white-noise conditions (reference = no noise, 0 dB and -10 dB SNR) were produced by 38 German native speakers (12m/26f). We measured acoustic features extracted from our target syllables (duration, intensity (average and range) and median fundamental frequency) and from the vowel within the syllable (F1, F2, and F2-F1 distance). We tested whether an unreliable noisy channel will trigger additive or interactive effects on speech enhancement arising from surprisal. We expected that Lombard speech and surprisal would interactively affect our acoustic variables, because speakers will enhance high surprisal units more than low surprisal units in noisy conditions. This strategy would allow speakers to be both informative and efficient.

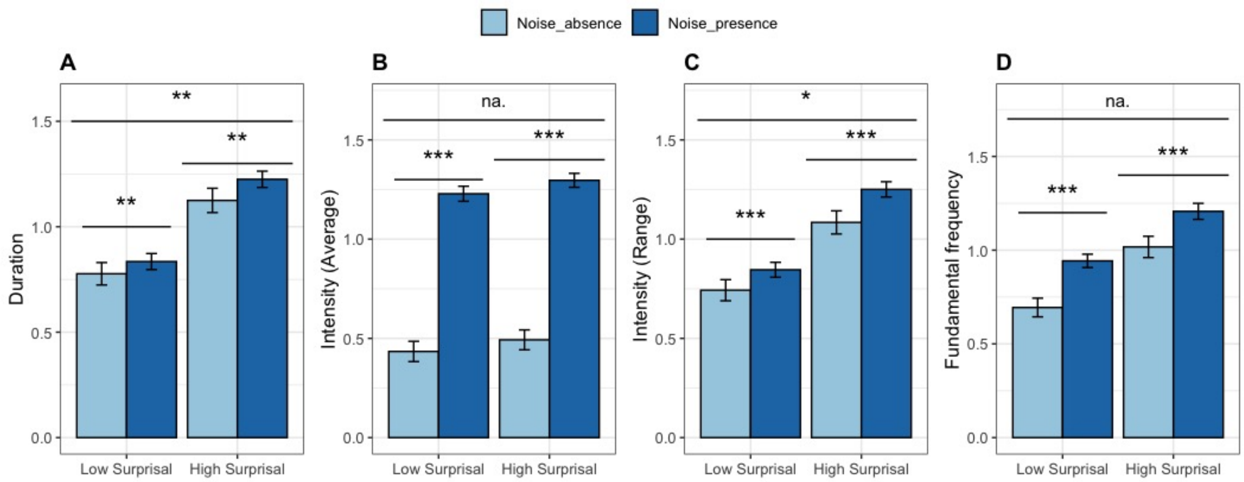

Figure 1: Mean (standard error) values of syllable-level features: (A) duration, (B) average intensity, (C) intensity range and (D) fundamental frequency as a function of absence vs. presence of noise (all values z-scored by speaker).

As expected from the Lombard speech literature, syllable duration was longer, intensity average was higher, intensity range was wider, and overall F0 was higher, when noise was present. However, the effect of noise levels was significant only for intensity (average and range) and fundamental frequency (F0). Unexpectedly, noise levels did not affect syllable duration. The effect of surprisal was significant only for syllable duration and intensity range, with longer duration and larger intensity range for high surprisal syllable (Fig. 1). The effect of absence vs. presence of noise reached statistical significance for F1, F2 and F2-F1 in front vowels. Front vowels exhibited higher F1, lower F2 and smaller F2-F1 in the presence of noise. However, the effect of presence of noise was only significant for F1 and F2-F1 in central and back vowels, with higher F1 and smaller F2-F1. No significant fixed effects of surprisal or noise levels were observed.

These findings suggest that speakers make an effort to increase the difference between syllables in high vs. low surprisal contexts in the presence of noise. Moreover, Lombard speech and predictability have additive effects in increasing syllable duration and intensity range, suggesting that properties pertaining to the channel are expanded without compromising source coding. No interaction was found between noise and predictability (Ibrahim et al. 2022b). This suggests that noise-related modifications might be independent of predictability-related changes, with implications for including channel-based and message-based formulations in speech production models. Although channel coding is not part of linguistic representation (message formulation) during planning, it shapes the phonetic output. Our study only explored one type of channel. Different types of channel abound, e.g., talking to L2 learners or robots, etc. Accounting for channel characteristics implies the need to consider communicative contexts as part of an enriched formulation of phonetic output during planning.

3.3. Effects of surprisal and prosodic structure on phrase final syllable duration. As prosody has been claimed to modulate information-theoretic effects on phonetic en-

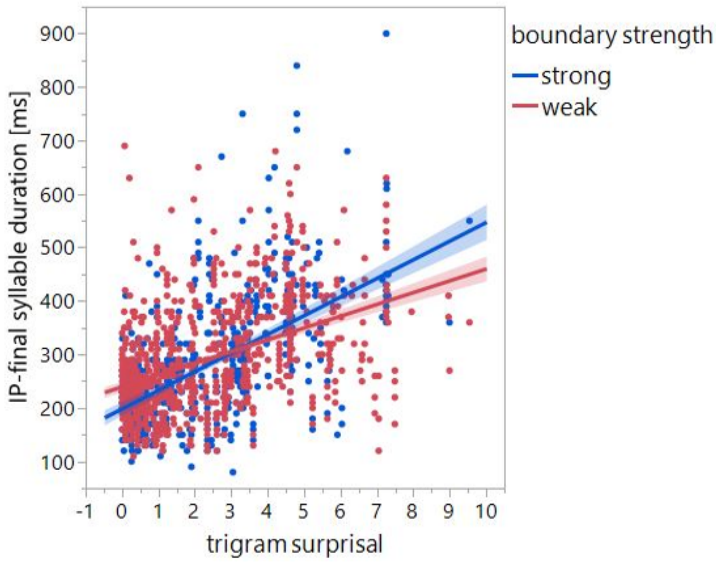

Figure 2: The effect of surprisal and boundary strength on IP final syllable duration.

coding, the possible interaction between levels of the prosodic hierarchy and the surprisal profile of utterances was investigated. This allowed us to assess the following research question: Is the surprisal effect fully attributable to prosody, as posited by the Smooth Signal Redundancy (SSR) hypothesis (Aylett & Turk 2004, 2006, Turk 2010) in its strong interpretation, or does surprisal remain effective in addition to prosody, as suggested for instance by (Baker & Bradlow 2009), or in some form of interaction between surprisal and prosody?

We examined the effects of prosodic structure (pitch accents and boundary strength) and surprisal on the duration of the final syllable in the intonational phrase. Phrase-final syllable durations and following pause durations were measured for a subset of a German radio-news corpus (DIRNDL, Börkelund et al. 2014), consisting of about 5 hours of manually annotated speech. The prosodic annotation is in accordance with the autosegmental intonation model and includes labels for pitch accents and boundary tones. We treated pause duration as a quantitative proxy for boundary strength. Surprisal was calculated as the surprisal of the syllable trigram of the preceding context, based on language models (LMs) trained on the DeWaC corpus. Syllable duration was statistically modelled as a function of surprisal and prosodic factors (pitch accent and boundary strength) in linear mixed effects models. We found a significant positive correlation between surprisal and phrase-final syllable duration. The results also revealed an interaction of surprisal and boundary strength in the final syllable of the intonational phrase. Syllable duration becomes longer with increasing surprisal, but this increase in duration is significantly greater before strong than before weak boundaries (Fig. 2). This modulation of pre-boundary syllable duration was observed above and beyond the well-established phrase-final lengthening effect (Andreeva et al. 2020).

Overall, our results show an interaction of surprisal and boundary strength , indicating that these factors complement each other in explaining syllable duration variability (Andreeva et al. 2020). Surprisal has a larger effect size than prosodic factors, indicating that the modulation between surprisal and phonetic encoding by means of prosody is not comprehensive. The prosodic

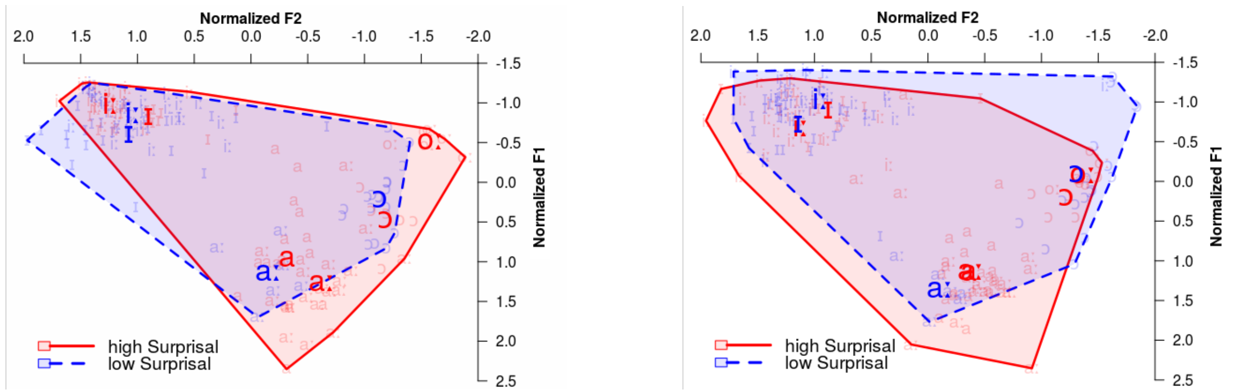

Figure 3: Vowel space of Bulgarian L2 speakers under high and low surprisal at advanced (left) and intermediate (right) proficiency level.

structure is not implementing much of the significant effects of redundancy, supporting the weak version of the SSR hypothesis (Aylett & Turk 2004, 2006).

3.4. Surprisal as a measure of language proficiency. This study started from the assumption that native speakers of a language share the same LM, with some degree of individual variability due to idiolectal, sociolinguistic or regional factors. Due to their exposure to the L2, language learners presumably build mental models of the predictability of linguistic events in their target language. These models will vary as a function of the speaker's proficiency level and amount of exposure. We investigated if surprisal factors of the target language (German) can explain phonetic variability, and here specifically vowel dispersion, of Bulgarian L2 speakers at different proficiency levels. Our analysis thus introduces a new approach to investigating language learning from an information-theoretic perspective. We predicted that the relation between surprisal and patterns of vowel dispersion observed in L1 speakers is also apparent in advanced proficiency level (C2) language learners, but less pronounced, or even non-existent, in intermediate proficiency level (B2) learners.

Vowel dispersion was measured for stressed tense and lax vowels in read speech from L1 German speakers, and advanced (C1) and intermediate (B2) Bulgarian speakers of L2 German. We calculated Pearson's r correlations between vowel dispersion and surprisal per speaker group. Vowel dispersion and triphone of the preceding context were significantly correlated for the L1 speakers and the L2 speakers at the C2 level. There was no significant correlation for the B2 speakers (Brandt et al. 2019). We then calculated three different LMMs for each speaker group with surprisal, word class, average vowel duration, and vowel tenseness as fixed factors, and speaker and word as random factors. The results revealed that Bulgarian L2 speakers, in particular at the B2 level, showed more vowel dispersion than German L1 speakers. Moreover, although all analyzed vowels were stressed, the B2 speakers raised their vowel space under low surprisal, thus reflecting Bulgarian L1 vowel raising in unstressed vowels (Andreeva et al. 2013) (Fig. 3). This can be interpreted as a certain degree of awareness in Bulgarian B2 speakers of the German phonological structures and their predictabilities. But they were not able to produce the target-language reduction pattern for vowels in low surprisal context and instead relied on their L1 reduction pattern.

4. Discussion. Phonetic structures expand temporally and spectrally when they are difficult to predict from their context. To some extent, effects of predictability are modulated by prosodic structure. Our research started from the assumption that speakers modulate details of the phonetic signal in the service of maintaining a balance of the complementary relation between surprisal (structural information density) and phonetic encoding. In this paper we have reported the results of experiments and corpus data analyses that were designed to test this hypothesis. We analysed the effects of surprisal on phonetic encoding, in particular on dynamic vowel formant trajectories, stop consonant voicing, syllable duration, and vowel space size. We controlled for several factors related to the prosodic structure, specifically lexical stress and major prosodic boundaries, as control factors in the statistical models that accounted for phonetic effects of changes in surprisal (e.g. Malisz et al. 2018, Brandt et al. 2021). Moreover, we investigated possible interactions between surprisal and speech enhancement in noisy conditions (the well-known Lombard effect) and between surprisal and the prosodic structure.

In what follows we discuss the primary results of the individual analyses and experiments. The investigation of the impact of predictability and prominence on the dynamic structure of the first and second formants of German vowels (section 3.1.1) revealed that effects of contextual predictability on fine phonetic detail can be observed not only in pointwise measures but also in dynamic features of phonetic segments. Surprisal based on the following context significantly explained the formant trajectory shape in German. This result is not necessarily expected since we also know from previous work that the direction of the surprisal effect depends on which acoustic measure is investigated. For instance, segment duration can be explained by surprisal of the preceding and following contexts, whereas vowel dispersion is only predicted by surprisal of the preceding context (Malisz et al. 2018).

Our analysis of the effect of syllable-based predictability on the durational correlates of the phonological stop voicing contrast in German (section 3.1.2) indicated an interaction effect of predictability and the voicing status of the target consonants on voice onset time, but a uniform effect on closure duration. This finding supports the interpretation that the effect of predictability on a phonetic cue is sensitive to this cue's phonological relevance in the specific language.

Findings from our study on the interaction of the effects of surprisal and speech enhancements in noise (section 3.2) support the interpretation that speakers make an effort to increase the difference between syllables in high vs. low surprisal contexts in the presence of noise. Apparently, Lombard speech and unpredictability have additive effects in increasing syllable duration and intensity range, suggesting that channel-related properties are expanded without compromising source coding. As no interaction was found between noise and surprisal, noise-related modifications may be independent of those induced by surprisal. If so, speech production models should include channel-based as well as message-based formulations: although channel coding is not part of linguistic representation (message formulation) during speech planning, it does shape the phonetic output.

Our case study on surprisal as an indicator of language proficiency (section 3.4), based on read speech productions by advanced and intermediate Bulgarian speakers of L2 German, with L1 German speakers as a control group, revealed a certain degree of awareness in intermediate L2 speakers of the German phonological structures and their predictabilities, whereas they failed to produce the target-language reduction pattern for vowels in low surprisal context and instead relied on their L1 reduction pattern. In contrast, advanced L2 speakers' vowel spaces showed dispersion patterns more similar to those of L1 German speakers. This finding suggests that language learners build mental models of the predictability of linguistic events in their target language with increased exposure to, and proficiency in, the L2.

Finally, we examined the effect of elements of the prosodic structure, viz. pitch accenting and boundary strength, and surprisal on the duration of the final syllable in the intonational phrase (section 3.3). The results revealed an interaction of surprisal and boundary strength in the final syllable of the intonational phrase: syllable duration becomes longer with increasing surprisal, and this effect was significantly enhanced before strong than before weak boundaries, above and beyond the well-established phrase-final lengthening effect. We thus found evidence for an interaction with surprisal and boundary strength. These factors complement each other in explaining syllable duration variability. As surprisal had a larger effect size than prosodic factors and the prosodic structure does not seem to implement much of the significant effects of redundancy, the modulation between surprisal and phonetic encoding by means of prosody is arguably not comprehensive. Therefore, our findings are generally compatible with a weak version of the Smooth Signal Redundancy hypothesis (Aylett & Turk 2004, 2006, Turk 2010).

Taken together, our findings suggest that the prosodic structure mediates between requirements of efficient communication and the speech signal. However, this mediation is not perfect: we found evidence for additional, direct effects of changes in predictability on the phonetic structure of utterances. Moreover, these effects appear to be stable across different speech rates in models that were fit to data derived from six different European languages (Malisz et al. 2018).

Funding. This research was funded in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project-ID 232722074 – SFB 1102 (project C1).

References

Andreeva, Bistra, William J. Barry & Jacques Koreman. 2013. The Bulgarian stressed and unstressed vowel system. a corpus study. In *Proceedings of Interspeech 2013 (Lyon, France)*, 345–348. https://doi.org/10.21437/Interspeech.2013-97.

Andreeva, Bistra, Bernd Möbius & James Whang. 2020. Effects of surprisal and boundary strength on phrase-final lengthening. In *Proceedings of Speech Prosody 2020 (Tokyo, Japan)*, 146–150. http://dx.doi.org/10.21437/SpeechProsody.2020-30.

Aylett, Matthew & Alice Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47. 31–56. https://doi.org/10.1177/00238309040470010201.

Aylett, Matthew & Alice Turk. 2006. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America* 119(1). 30–48. https://doi.org/10.1121/1.2188331.

Baker, Rachel A. & Ann R. Bradlow. 2009. Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech* 52(4). 391–413. https://doi.org/10.1177/0023830909336575.

Baroni, Marco, S. Bernardini, A. Ferraresi & E. Zanchetta. 2009. The Wacky Wide Web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43. 209–226. https://doi.org/10.1007/s10579-009-9081-4.

Börkelund, Andres, Kerstin Eckart, Arndt Riester, Nadja Schauffler & Katrin Schweitzer. 2014. The extended DIRNDL corpus as a resource for automatic coreference and bridging resolution. In *Proceedings of LREC (Reykjavik)*, 3222–3228.

Brandt, Erika, Bistra Andreeva & Bernd Möbius. 2019. Information density and vowel dispersion in the productions of Bulgarian L2 speakers of German. In *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS 2019) (Melbourne, Australia)*, 3165–3169.

Brandt, Erika, Bernd Möbius & Bistra Andreeva. 2021. Dynamic formant trajectories in German read speech: Impact of predictability and prominence. *Frontiers in Communication / Language Sciences* 6. 1–15 (643528). https://doi.org/10.3389/fcomm.2021.643528.

Chan, D., A. Fourcin, D. Gibbon, B. Granström, M. Huckvale, G. Kokkinakis, K. Kvale, L. Lamel, B. Lindberg, A. Moreno, J. Mouropoulos, F. Senia, I. Trancoso, C. Veld & J. Zeiliger. 1995. EUROM - a spoken language resource for the EU - the SAM projects. In *Proceedings of Eurospeech (Madrid, Spain)*, 867–870. https://doi.org/10.21437/Eurospeech.1995-198.

Cooke, Martin, Simon King, Maeva Garnier & Vincent Aubanel. 2014. The listening talker: a review of human and algorithmic context-induced modifications of speech. *Computer Speech and Language* 28(2). 543–571. https://doi.org/10.1016/j.csl.2013.08.003.

Eckart, Kerstin, Arndt Riester & Katrin Schweitzer. 2012. A discourse information radio news database for linguistic analysis. In Christian Chiarcos, Sebastian Nordhoff & Sebastian Hellmann (eds.), *Linked data in linguistics: Representing and connecting language data and language metadata*, 65–75. Springer.

Hale, John. 2016. Information-theoretical complexity metrics. *Language and Linguistics Compass* 1–16. https://doi.org/10.1111/lnc4.12196.

Ibrahim, Omnia, Ivan Yuen, Bistra Andreeva & Bernd Möbius. 2022a. The effect of predictability on German stop voicing is phonologically selective. In *Proceedings of Speech Prosody 2022 (Lisbon, Portugal)*, 669–673. https://doi.org/10.21437/SpeechProsody.2022-136.

Ibrahim, Omnia, Ivan Yuen, Marjolein van Os, Bistra Andreeva & Bernd Möbius. 2022b. The combined effects of contextual predictability and noise on the acoustic realisation of german syllables. *Journal of the Acoustical Society of America* 152(2). https://doi.org/10.1121/10.0013413.

Levy, Roger & T. Florian Jaeger. 2006. Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt & T. Hoffman (eds.), *Advances in neural information processing systems 19*, 849–856. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/7503.003.0111.

Malisz, Zofia, Erika Brandt, Bernd Möbius, Yoon Mi Oh & Bistra Andreeva. 2018. Dimensions of segmental variability: interaction of prosody and surprisal in six languages. *Frontiers in Communication / Language Sciences* 3. 1–18 (article 25). https://doi.org/10.3389/fcomm.2018.00025.

Mayer, Jörg. 1995. Transcription of German intonation—the Stuttgart system. Tech. rep. Institute of Natural Language Processing, University of Stuttgart.

Möhler, Gregor, Antje Schweitzer, Breitenbücher & Martin Barbisch. 2000. IMS German Festival (version 1.2-os). University of Stuttgart: Institut für Maschinelle Sprachverarbeitung (IMS). Retrieved 2020-01-02.

Schiel, Florian. 1997. Siemens Synthesis Corpus – SI1000P. University of Munich.

Stolcke, Andreas. 2002. SRILM—an extensible language modeling toolkit. In *Proceedings of Interspeech 2002 (Denver, CO, USA)*, 901–904.

Turk, Alice. 2010. Does prosodic constituency signal relative predictability? A Smooth Signal Redundancy hypothesis. *Laboratory Phonology* 1. 227–262. https://doi.org/10.1515/LABPHON.2010.012.

Witten, I. H. & T. C. Bell. 1991. The zero-frequency problem: Estimating the probabilities of novel events in adaptive text compression. *IEEE Transactions on Information Theory* 37(4). 1085–1094.