EXPANDING FIELD STUDIES USING ONLINE SPEECH PERCEPTION EXPERIMENTS

Amelia E. Kimball[1], Hermann Keupdjio[2], Kathryn Franich[3], Constantine Kouankem[4]

Laboratoire de Linguistique Formelle[1],CNRS[1],Université de Paris Diderot [1], Sorbonne Paris Cité[1] ,University of British Columbia[2] ,University of Delaware[3], University of Yaoundé 1[4]
akimball@linguist.univ-paris-diderot.fr, keupsmann2011@gmail.com, kfranich@udel.edu, kconsty2000@yahoo.fr

ABSTRACT

Collecting data from understudied languages is a vital enterprise that enriches our knowledge of the nature of human language. Accomplishing this with in-person visits is invaluable, but there is an urgent need for more data, and a limited number of linguists with the training and resources to conduct field work. In this situation, online experiments provide a powerful supplementary tool for linguists and fieldworkers studying underdocumented languages. Specifically, we show how, rather than supplanting fieldwork, online experiments allow for expansion of field work with pre-visit pilots and follow-up experiments. We argue they are a helpful tool in creating and enhancing global collaborations between field linguists, members of understudied language communities, and linguists without field training. We also provide solutions to common challenges of implementing online experiments on under-studied languages. We show as an example our experiment on "stress deafness," implemented in Medumba, a language spoken in Cameroon.

Keywords: Underdocumented Languages, online methodologies, speech perception, Medumba.

1. INTRODUCTION

It is in the interest of all linguists to build their conception of human language based on data from a wide variety of languages, but data and scholarship remains concentrated on a small number of well-studied languages. A major barrier to having equality of representation among the world's languages is a lack of resources: traveling to and staying in places where underdocumented languages are spoken for the duration of time necessary to conduct research can often be expensive, even if the linguist is native to or resides in the country where the language is spoken. One solution to this problem is to supplement in-person visits with online experiments. Compared to an in-person visit, online experiments are faster to implement, less expensive, and more scalable. Let us be perfectly clear: in our view online experiments *are no substitute* for a visit from a trained field linguist. We nonetheless believe they are an invaluable tool for researching understudied languages for four reasons.

First, online experiments allow for expanding the timeline of a field study. Rather than concentrating experiment development and data collection during a field visit, linguists can conduct pilot studies before a trip, fine-tuning their materials so as to make their time in the field maximally productive. Online data collection can also be used for follow-up studies after a trip if difficult circumstances arise and data collection is not completed, or additional measures are needed. Online experiments also provide a source of pilot data for grant proposals for linguists interested in conducting future in-person fieldwork.

Second, online experiments are easily spread through email, text, or social media, and allow linguists to connect with many more speakers, or connect from afar. A link to a survey can be distributed, and participants can then easily forward the link to friends or relatives. Furthermore, using an online interface it is possible to run many subjects simultaneously, either in person (e.g., a group of participants in an internet café) or remotely.

Third, because building online experiments is easily taught and requires relatively few resources, it allows for the participation of a larger population of researchers. For example, many open source survey-building platforms come with easily accessible, highly detailed instructions for those with even minimal experience. Moreover, online methods may promote collaboration and pooling of resources among linguists of different backgrounds with varying access to training opportunities. The ability to create and edit surveys from anywhere in the world means that research partners from different institutions across the globe can help one another to fine-tune experimental design. This is useful in the case of understudied languages, as access to knowledge and training may be a challenge for local communities where such languages are spoken.

Lastly, one powerful reason these experiments are helpful is that by simplifying data collection, we can begin to collect substantial data on a more diverse range of languages. In other words, this type of data collection is simple enough that a "sample of convenience" need not be restricted to western

university students, but instead can attempt to reflect the diverse languages of the world.

2. IMPLEMENTING AN ONLINE EXPERIMENT IN A FIELD SETTING

2.1. Identifying possible studies for online experiments.

It is clear that many types of data collection are not possible remotely. Any study which must involve a researcher (e.g., a study with equipment that requires training, a study that involves interaction with the researcher) will not be possible to implement. Geographical restrictions exist too, as areas where internet access is scarce or expensive will not be ideal candidates for an online experiment. However, there are many research areas where internet use is widespread, and there is a subset of research questions that can be addressed with data collected from questionnaires. As a general rule, anything that can be formatted as a questionnaire can be implemented as an online survey (e.g. grammaticality/acceptability judgements, cloze tests, garden path reading tests). When audio files are added as part of the survey, this capacity is increased, as listeners are able to make judgements on heard speech, and so can perform many linguistic tasks targeting speech perception (a categorical perception task, matched guise task, judgements of tone and intonation, working memory tasks, etc.).

2.2. Implementing the experiment itself

Creating an experiment is as easy as creating a short online survey. Survey platforms (e.g. Qualtrics, Googleforms, SurveyMonkey, Typeform), typically provide the ability to know where a participant participated from, how long each question took, and immediate feedback based on participant answers, as well as step-by-step instructions on how to build and experiment. We will focus here on providing general information on some of the most common concerns linguists who are new to online experiments (or new to implementing them in field locations) might have.

2.3. Common concerns

2.3.1. What is the value of experimental data for the field linguist?

We do not claim that experiments are better than other methods of research, or that all field linguists should become experimentalists, but for many linguists there is value in using an experiment to test a narrow question in a controlled way as part of a larger study. For example, field linguistics as a course is often taught using one or two informants from a language. An online experiment could allow students in such a class to get judgements and input from more speakers, possibly highlighting points of dialect variation which allow for a more nuanced linguistic analysis. Another particular strength of experiments is in collecting perception measures or judgements on finely controlled stimuli, which is not possible in an interview setting or in typical conversation.

2.3.2. Creating stimuli and instructions

Creating stimuli in underdocumented languages may be challenging or impossible if no native speakers are available to record or write stimuli. Audio recording, may help in some cases. Platforms such as JotForm enable remote audio recording without the need for a speaker to have recording equipment of their own (other than a computer). Care must also be taken in the development of experiment instructions such that they are easily interpretable by study participants. This is particularly true for languages which are not written or use orthographies that are not widely taught. In such cases, pictures of the orthography may be used where type symbols are not displayed properly, or a lingua franca can be used for written instructions. Alternatively, it is possible to develop auditory instructions (i.e. a set of recordings of the instructions that participants can play to themselves), rather than written ones. This is a useful option in situations where speakers may not be able to read in the language under study or a lingua franca.

2.3.3. Paying participants

Paying participants may be difficult depending on the institutional funding source, but we have found it to be possible in the form of an honorarium given to a local assistant, who further distributed funds to other participants. Amazon Mechanical Turk (AMT) is another option for payment. AMT provides an online bulletin board of tasks which workers can sign up to do [for more information see 1,2]. The workers may be located anywhere in the world, and so participants could be paid through this interface if funding regulations allow it. Of course, it is important to make sure participants are paid enough to cover their time, and it may also be necessary to compensate for smartphone data usage depending on the study.

2.3.4. Research permits, institutional approval, and data storage

As with in-person data collection, it is often necessary for a researcher to obtain permission to conduct their work from local government, especially if they are not from the community where the language is

spoken. While it may be easier to conduct web-based research without such permissions, we stress that web-based research should be treated no differently from in-person data collection as far as such ethical issues are concerned.

For individuals collecting data outside of their home country, we have found that an institutional review board may require additional information about the local situation in order to approve the experimental setup. As with any sort of international project, institutional review boards will require translated consent forms and subject recruitment language (e.g. social media posts or emails used for recruitment) and may also require the completion of a certificate in creating ethical online experiments.

Of course, care must be taken in how data is stored to protect personally identifiable information. This is easily done in an online study, because data can be stored in a remote server at the home institution. This also removes worries about backing up or losing data during a field visit.

2.3.5. Technical problems

Software exists to create complex experiments online (e.g. Psychopy, JsPsych, Psytoolkit, Testable, Labvanced, Pebl, Finding Five) and linguists with programming experience may be able to program their own experiments. However, coding an experiment that will work on any device, on any browser, is no small task. For cases where a researcher is not present to provide technical help, it may be advantageous to choose one of the many survey programs available rather than programming an experiment. It is crucial to test the experiment in the local setting to make sure that it loads correctly.

3. EXAMPLE EXPERIMENT: STRESS DEAFNESS IN SPEAKERS OF MEDʉMBA

3.1. The Current Experiment

Our experiment investigated the phenomenon of 'stress deafness', or the tendency for speakers of some languages to be unable to report the location of word stress in a second language. The main theoretical issue lies in identifying what it is that makes these speakers "stress deaf." Past experiments have suggested that the presence or predictability of lexical stress in the L1 are key factors [3,4]

The classic experiments on stress deafness test French speakers on their ability to locate stress in English words, and French speakers are found to be "stress deaf." Medʉmba speakers make an interesting test case because unlike French, Medʉmba has no acoustic evidence of stress [5]. However, as a tonal language Medʉmba uses two key markers of stress in English, f0 and duration [6] to cue tonal patterns. Tone in Medʉmba is also 'unrestricted', in that neither high nor low tones (which make up the Medʉmba tonal inventory) must occur in specific positions within a word. Thus, tone assignment has the same sort of unpredictability that stress assignment has in languages like English or Dutch. Different from stress-based languages, however, is the fact that high tones in Medʉmba are associated with shorter duration, not longer duration [6,7,8].

If Medʉmba speakers are "stress deaf", in a word discrimination and memory task we predict high performance for minimal pairs of words that differ by phoneme, but low ability to discriminate minimal pairs that differ by stress location, and even lower ability if the pitch cue to stress is removed. If instead Medʉmba speakers are sensitive to stress, we should see above chance performance across all conditions.

3.2. Situation of Medʉmba

Medʉmba is Grassfields Bamileke Bantu language spoken in (Francophone) Western Cameroon. There are an estimated 210,000 speakers, many of whom are multilingual (typically in French and one or more other Cameroonian mother tongues). Cameroon — along with Nigeria, Sudan, and Ethiopia — is reported to have one of the highest language mortality rates in Africa [9,10]. Internet and smartphone use is widespread and growing in Cameroon, with one estimate of the internet penetration rate at 35%. [11] Approximately 75% of Cameroonians have a mobile phone, with 40% of those being smartphones. [11]

3.3. Method

3.3.1 Stimuli

The procedure followed [3], Experiment 3, though it is not an exact replication. Participants are presented with disyllabic noncewords differing in either a segmental phoneme or in the location of stress and asked to identify which word they've heard by pushing a button they've been trained to associate with that word. The task included three conditions: 1) the PHONEME condition, in which participants chose between the words /miga/ and /miba/, which differed only in the place of articulation of the third segment; 2) the STRESS condition, in which the choice was between KIga or kiGA, which differed in the placement of stress on either the initial or final syllable; 3) the STRESS-PITCH FLATTENED condition, where the same words from the stress condition were presented but with their pitch flattened to 150 Hz using PSOLA in Praat [12]. The words were recorded by a male native speaker of English and each uttered 6 times to incorporate some phonetic variability as

was present in the original study by [3]. For the stress condition, words were uttered both phrase-medially and phrase finally, in order to incorporate some level of pitch variability; three tokens from each prosodic position were used. All words were non-existent but phonotactically licit words in Medumba both from a segmental and tonal perspective. Recordings of words were concatenated before a recording of a different male speaker saying 'OK' to ensure that judgments did not rely on echoic memory [13].

3.3.2 Online study design

An online survey was built using Qualtrics survey software. Participants were recruited by email, phone and social media using connections formed by previous research and personal connections to the language community. Participants first read a consent form and consented to the experiment. For the first block, the phoneme block, they were given eight two-word practice trials with feedback. They had to answer correctly on all training trials before moving to the main task within the experimental block. After training, they proceeded through three sets of trials, with first two, then three, then four word sequences consisting of a mix of the words they were trained on (e.g. the sequence *miga miba miba* for a three-word trial). This procedure (including training) was then repeated twice, once for the stress condition and once for the stress-pitch flattened condition. We report data from 16 participants from towns near Bangangté, where Medumba is primarily spoken.

3.3. Results

Results are shown below in fig. 1. Data were analysed using mixed effects logistic regression in the lme4 package for *R* [14]. The model included fixed effects of condition (3 levels) and sequence length (3 levels) and their interaction, as well as by-subject random slopes for each factor. These categorical variables were both sum-coded. There were significantly more errors as the stimulus length increased (β = .808; t = 7.79; p < 0.001). Furthermore, there were significantly more errors made in the stress condition than in the phoneme condition (β = .285; t = 2.02; p < 0.05), and even more errors made in the stress-pitch flattened condition than in the plain stress condition (β = .285; t = 2.01; p < 0.05). Error rates in the two word stress condition averaged 11%, and then jumped to 28% for 3 word sequences, and 43% for 4 word sequences. Comparing our results from the stress condition to those from [3], Experiment 3, error rates were more comparable to those found for speakers of French (29%; 28%; 59%), who are labelled 'stress deaf' than for speakers of Spanish (0%; 4%; 10%), who are not. This holds even when controlling for the slightly higher overall error rate in our experiment.

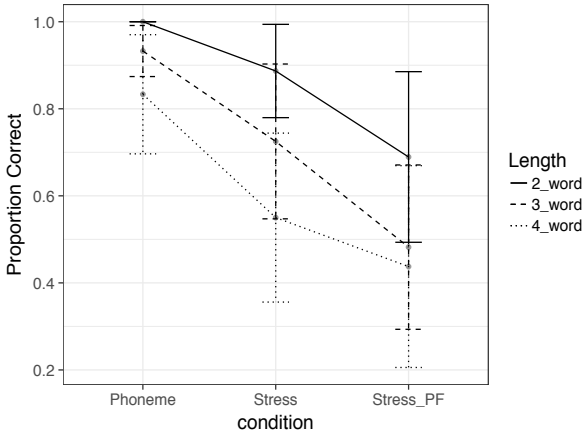

Figure 1: Proportion of correct responses by stimulus length and condition. Error bars represent 95% confidence intervals.

3.3. Conclusions of our experiment.

Overall, our data show that Medumba speakers are less good at detecting words that differ in stress compared to words that differ in segmental phonemes, and that removal of pitch variation poses additional difficulties in stress perception and memory. Compared with results from prior studies on French and Spanish speakers, Medumba speakers show an intermediate pattern of stress deafness, but a pattern closer to 'stress-deaf' French speakers than to Spanish speakers. Of course, there were some differences in the stimuli used in previous studies, so we will need to examine how speakers of languages like French and Spanish do on our experiment before drawing any strong conclusions. Thus far, however, our results support earlier findings that the phenomenon of stress deafness is linked not only with (un)predictability of the location of lexical stress and lexical tone [3,4], but also with the acoustic correlates used to cue stress or tone in a language [15].

4. GENERAL DISCUSSION

We highly recommend online experiments as a tool for researching understudied languages. We believe they are an excellent tool for creating broader collaborations and collecting more data with fewer resources. The implementation of our experiment has enabled us to connect with potential future informants, and plan future research. Our results also challenge the neat classification of speakers into categories of stress-deaf or not stress-deaf based on previous studies. This supports one of our goals as field linguists, namely to ensure that theories of language account for the variability in the world's languages.

Acknowledgements: AK was supported by French Agence Nationale de la Recherche grants ANR-17-CE28-0009 (GEOMPHON) and ANR-10-LABX-0083 (EFL)

7. REFERENCES

[1] Author 1, manuscript on building experiments in Mechanical Turk.

[2]Paolacci, G., Chandler,M., Iperotis,P., 2010 Running Experiments on Amazon Mechanical Turk, *Judgement and Decision Making* 5,411-419

[3] Dupoux,E.,Peperkamp,S.,Sebastián-Gallés,N, A Robust Method to study stress "deafness" *J. Acoust. Soc. Am, 110 (3) 1606-1618*

[4] Rahmani, H., Rietveld, T., and Gussenhoven, C. (2015). Stress "deafness" reveals absence of lexical marking of stress or tone in the adult grammar. *PLoS One*, *10*(12), e0143968.

[5] Franich, K. (2017). *The Interaction of Prominence, Rhythm, and Tone in Medumba*. UChicago Dissertation.

[6] Franich, K. (2016). Internal and contextual cues to tone perception in Medumba. *J. Acoust. Soc. Am Exp. Let.*, *140, EL107*.

[7] Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *JASA 27*, 765–8.

[8] Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech 1*, 120–52.

[9]Okal, Bernad Odoyo (2014). "Benefits of Multilingualism in Education". *Universal Journal of Educational Research*. **2** (3): 228.

[10] Opala, K. (6 March 2002). "How Kenya Stands to Lose More than 10 Tribes". *Daily Nation*. Nairobi: Nation Media Group.

[11] Agence de Presse Africaine 2018 http://mobile.apanews.net/index.php/en/news/cameroon-internet-user-population-reaches-88-million accessed 09/12/2018

[12]Boersma,P.,Weenink,D, 2018, Praat: doing phonetics by computer [Computer program]. Version 6.0.43, retrieved from http://www.praat.org/

[13] Morton, J., Crowder, R., and Prussin, H.1971. Experiments with the stimulus suffix effect, *J. Exp. Psychol.* 91, 169–190

[14] Bates,D.,Martin,M.,Bolker,B.,Walker,S.,2015. Fitting Linear Mixed-Effects Models Using {lme4} *Journal of Statistical software*, 67, 1-48

[15] Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *JSLHR*, *57*(4), 1468-79.