Phonetic and Phonological Variation in Vowel Discrimination Performance: Effect of Swedish Vowel Categories and Dialects

Renata Kochančikaitė[1]*, Mikael Roll*[1]
[1] *Centre for Languages and Literature, Lund University, Sweden*
renata.kochancikaite@ling.lu.se, mikael.roll@ling.lu.se

Abstract

Acoustic discrimination of speech sounds is affected by various factors, ranging from more universal acoustic properties of categories to the phoneme systems of the native language and dialect, and even influences from languages learned later in life. A discrimination experiment containing East Central Swedish vowels was carried out with 30 native Swedish listeners in order to explore the variation in vowel discrimination performance. Both phonetic and phonological variables have been found to have an effect on discrimination performance. Peripheral location of vowels in the F1/F2 vowel space was found to increase the discrimination performance. South Swedish dialectal area was associated with a decreased discrimination performance. Continuous exposure to foreign languages other than English was not a significant factor.

Introduction

Various factors have been shown to affect the discrimination of speech sounds in perception. These factors include the listeners' sense of pitch (Sun et al., 2017), location of category boundaries in the listeners' native language (Liberman et al., 1957) and dialect (Escudero & Williams, 2012), as well as the acoustic salience of the phoneme categories themselves (Polka & Bohn, 2011). This study explores how well native Swedish speakers discriminate vowel sounds and what factors may affect their performance. The aim is twofold, involving phonetic and phonological factors. First, we will measure how much the categories' locations on the F1/F2 vowel space and their categorical boundaries affect the discrimination performance in Swedish speakers. Second, we will check whether there are differences in discrimination performance that could be attributed to dialectal background or foreign language exposure.

Background

Categorical perception

Listeners' perception of speech sounds is not identical to their acoustic properties. As infants grow older and are exposed to language, perceptual attunement to native phoneme categories occurs, facilitating quick categorisation which is necessary for efficient speech processing. This process reduces the accuracy of acoustic discrimination of contrasts *within* a native phonemic category, while discrimination *across* the category boundaries is retained and, for some native contrasts, enhanced (Junge et al., 2019). A phenomenon called *categorical perception* emerges, which means that a difference of the same acoustic magnitude is less perceptible between variants of the same phoneme than between two different phonemes of a given language (Liberman et al., 1957).

Phonetic and phonological factors

Some language-independent features exist that make certain vowel contrasts easier to discriminate than others. One such feature is focalization, that is, the convergence of formants that are close neighbours in the spectrum of a vowel (Schwartz et al., 2005). Due to the anatomy of the human ear, spectrally close formants are perceived as a single formant, with a centre of gravity that is an average of the individual formants (Chistovich, 1985). When formants merge, their acoustic energy is focused in the narrower region of the spectrum, thereby boosting the salience of such sounds.

Another relevant acoustic feature is the relative location of the vowel in the F1/F2 vowel space. Pairings of similar vowel sounds are easier to distinguish if they include a vowel closer to the periphery of the vowel space compared to pairings that only involve more centrally located vowels (Polka & Bohn, 2003). However, phonetic attributes of acoustic salience can be overridden by experience with phoneme contrasts in the native language (Polka & Bohn, 2011). There is evidence that even late L2 phoneme systems may have an effect on acoustic perception of the native categories (Cabrelli et al., 2019; Gorba, 2019).

Sweden is a long and narrow country, and, naturally, the Swedish language has many dialects. The vowel systems of these dialects differ from Standard Swedish not only in exact placement on the F1/F2 vowel space or the amount of diphthongisation, but also in the number of phonemes (Leinonen, 2010). Therefore, it is reasonable to assume that the specific phoneme systems found in Swedish dialects may influence the discrimination performance of Swedish speakers.

Swedish dialects

Modern variants of the Swedish language spoken inside the country are often divided into six dialectal areas: *Sydsvenska* (South Swedish, including provinces of Skåne, Halland, Blekinge, and southern part of Småland), *Västmellandsvenska* (West Central Swedish, including Bohuslän, Västergötland, Östergötland, and southern part of Värmland), *Östmellansvenska* (East Central Swedish, including Närke, Västmanland, Uppland, and Södermanland), *Dalabergslagska* (Dala Mountain Swedish, including Dalarna, Gästrikland, and northern part of Värmland), *Norrländska* (Northern Swedish, including Härjedalen, Hälsingland, Medelpad, Jämtland, Ångermanland, Västerbotten, Norrbotten, and Lappland), and *Gotländska*, the dialect spoken on the island of Gotland (Bruce, 2010). Due to many similarities between West and East Central Swedish, these dialects are often considered together and refered to as Central Swedish, but in this study they are treated as two separate dialectal areas.

Method

Participants

Thirty native Swedish speakers (21 males, 8 females, 1 not specified, average age = 27.9 years, SD = 5.48) without hearing difficulties participated. Eleven participants were raised in geographical area of Östmellansvenska, nine of Västmellansvenska, seven of Sydsvenska, two of Norrländska, and one of Dalabergslagska regional dialect varieties. Nine participants reported that they often read texts in other foreign languages than English (English was reported by all participants). All participants declared having been raised monolingual but five participants noted that, in addition to Swedish, another language was used in their family. All participants were recruited via online participant recruitment service Prolific and were paid for their participation.

Experiment design

A 2-alternative forced-choice AX discrimination task was created in PsychoPy (Peirce et al., 2019). The experiment consisted of three blocks with 144 randomised unique stimuli per block. One stimulus, or one pairing, consisted of two 325 ms-long synthetic vowel sounds from similar areas of the F1/F2 vowel space. In the pairing, there was a 475 ms gap between the sounds. Inter-stimulus interval was set to 1 s. Six continua between eight long East Central Swedish vowels were represented in the stimuli (Figure 1). Each pairing included vowel sounds from just one continuum, either from the edge (near the prototype representative of the vowel category) or from the middle of the continuum. The acoustic distance between the sounds in the pairing varied from 0 to 0.375 Bark in 0.075 Bark steps, yielding six different magnitudes of acoustic distance: 0, 0.075, 0.15, 0.225, 0.3, and 0.375 Bark. For each magnitude, twenty-four different pairings were created—four for each continuum—two at the edge and two at the middle position. In total, there were 144 unique pairings.

Stimuli

Category prototype formant values were derived from recordings of a native East Central Swedish speaker. The recordings were three repetitions of 16 words (*kok*, *fot*, *tår*, *kåk*, *tak*, *par*, *pär*, *häl*, *hel*, *fet*, *fyr*, *syl*, *tut*, *fur*, *tör*, *lös*) pronounced in isolation. Unrounded front high vowel /i:/ was omitted due to the very short distance from this vowel to the nearest neighbouring long vowels /e:/ and /y:/ in the F1/F2 vowel space, and because F3 rather than F2 is involved in the acoustic distinction between /i:/ and /y:/. Vowels from the recordings were analysed in Praat and average F1-F4 measurements in Hz were taken for each vowel during its steady state. Then, a synthesized copy of a prototype vowel of each category (/u:/, /o:/, /ɑ:/, /ɛ:/, /e:/, /y:/, /ʉ̞:/, /ø:/) was recreated using the Soundgen package in R, based on the formant measurements obtained from the actual recordings. The timbre, intensity curve, and pitch curve were adjusted in Soundgen to match the typical pronunciation pattern of the recorded speaker. Other parameters—duration, mouth opening, temperature, length of the vocal tract, and formant width—were also adjusted so as to resemble the original recordings as much as possible. All parameters, except formant values, were kept the same across the different vowel types.

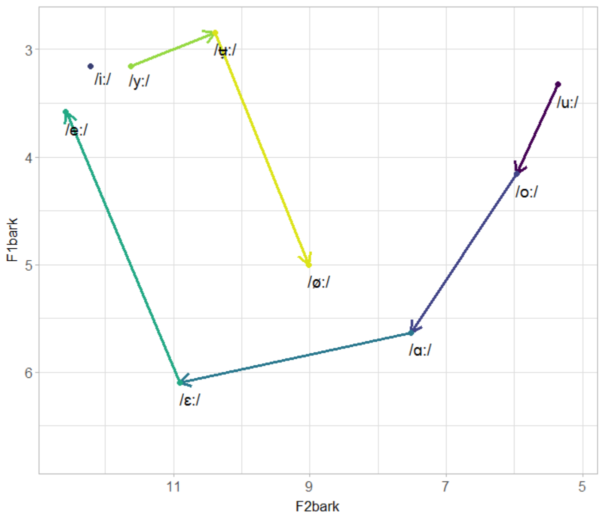

Figure 1. Approximate location of prototype vowels and direction of vectors along which the inter-category vowels were synthesised.

Six inter-category vectors in the F1/F2 vowel space were defined, connecting the prototype vowels in the following continua: /u:/ → /o:/, /o:/ → /ɑ:/, /ɑ:/ → /ɛ:/, /ɛ:/ → /e:/, /y:/ → /ʉ̞:/, /ʉ̞:/ → /ø:/. The lengths of these vectors were calculated in Manhattan distance, by subtracting the starting point formant value from the ending point formant value, then adding the ΔF1 and ΔF2 together and dividing the sum by 2. Calculations were carried out in Bark scale. Formant values to be used in the synthesis of the inter-category, non-prototype vowel sounds were taken from points on the inter-category vectors. These points were placed along every vector by dividing it into 0.075 Bark steps. Due to the different lengths of the inter-category continuum vectors, the number of intermediate vowel sounds generated per continuum differed (see Table 1).

Table 1. List of continua used in the experiment. The "Length" column shows the total number of intermediate vowel sounds generated for each inter-category vector.

Continuum	*Length*
/u:/ → /o:/	10
/o:/ → /ɑ:/	20
/ɑ:/ → /ɛ:/	24
/ɛ:/ → /e:/	26
/y:/ → /ʉ̞:/	12
/ʉ̞:/ → /ø:/	24

Procedure

The task was hosted on the online platform Pavlovia. The participants ran the experiment on their own desktop or laptop computers and completed it using their own earphones/headphones. The participants were instructed to listen carefully to the pairings and decide whether the two sounds were completely identical ('same') or somewhat different ('not same'). The participants responded after each stimulus presentation by pressing arrow keys on the keyboard: left for 'same' and right for 'not same'. After the task, they completed a survey where they indicated what Swedish city and/or province they were raised in, which dialects they heard at home during childhood, what languages they often read in, and whether there were other languages than Swedish used in their childhood home.

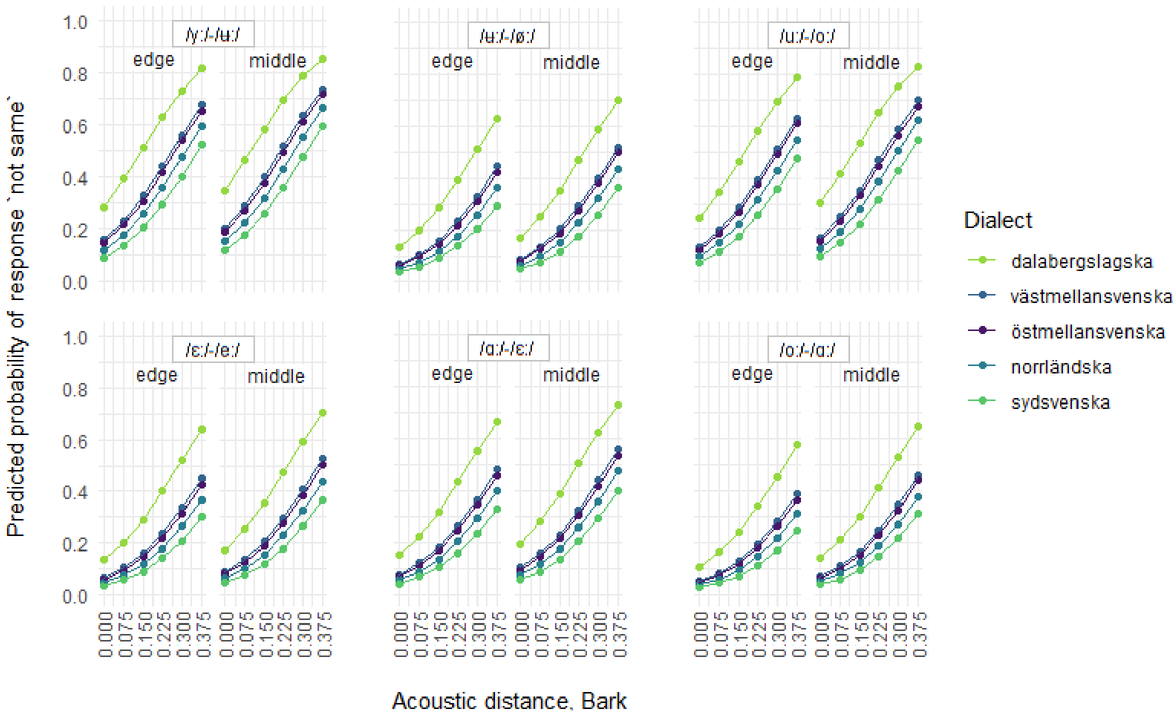

Figure 1. Results of the best-fitting logistic regression model. The increasing likelihood of discrimination is shown as a function of increasing acoustic distance, with different colour lines showing dialectal differences in the discrimination function. The results are shown in separate plots for better viewing. The plots are split by continuum and stimulus position on the edge (i.e. near a prototype category) versus in the middle of the inter-category continuum (i.e. on a category boundary).

Analysis

A logistic regression was performed to test the effects of acoustic distance, specific continua, position of the stimulus on categorical boundary in the middle of the continuum versus at the prototype edge, participants' main dialectal area, and exposure to foreign languages (other than English) on the likelihood that the participants would discriminate the sounds in the pairing and respond 'not same' versus 'same'. Östmellansvenska was chosen as the reference category for the dialectal area variable, due to its similarity to Standard Swedish. /ɑ:/ → /ɛ:/ continuum was allocated alphabetically as the reference category for continuum variable. AIC model selection was used to distinguish among a set of possible models.

Results

The best-fit logistic regression model included all parameters except exposure to foreign languages which showed no effect on the outcome. The predicted response likelihood is shown in Figure 2. The model was statistically significant ($AICc$ = 1.016, df = 11, $p < 0.001$) and explained 13% (McFadden's R^2) of the variance in responses. One unit (0.075 Bark) increase in acoustic distance was associated with a 62% increase in the likelihood that the participants would discriminate the sounds in the pairings (OR = 1.62, 95% CI [1.58, 1.67]).

Compared to the reference dialect area Östmellansvenska, one dialect area, Dalabergslagska (OR = 2.34, 95% CI [1.88, 2.91], $p < 0.001$, N = 1), was associated with a 134% increase in the likelihood of discrimination. Two dialect areas were associated with a significant decrease in the likelihood of discrimination: Sydsvenska (OR = 0.57, 95% CI [0.51, 0.65], $p < 0.001$, N = 7) and Norrländska (OR = 0.78, 95% CI [0.65, 0.93], p = 0.006, N = 2).

Compared to the reference continuum /ɑ:/ → /ɛ:/, two continua were associated with a significant increase in the likelihood of discrimination, /y:/ → /ʉ̟:/ (OR = 2.23, 95% CI [1.94, 2.57], $p < 0.001$) and /u:/ → /o:/ (OR = 1.82, 95% CI [1.58, 2.17], $p < 0.001$). Two continua were associated with a significant decrease in the likelihood of discrimination, /o:/ → /ɑ:/ (OR = 0.68, 95% CI [0.58, 0.79], $p < 0.001$) and /ʉ̟:/ → /ø:/ (OR = 0.85, 95% CI [0.73, 0.99], p = 0.035).

Compared to edge position, middle position of the pairing was associated with 35% increase in the likelihood of discrimination (OR = 1.35, 95% CI [1.24, 1.47]).

Discussion

The discrimination function depends on both phonetic and phonological factors. Acoustic features of the speech sounds define their perceptual salience. At the same time, differences between phoneme systems of the various dialects move the category boundaries and thereby focus individual attunement to different hotspots in the acoustic space. This study has investigated these factors and found significant effects in both of them.

We observed that discrimination performance varies depending on which inter-category continuum the pairing is taken from. Among the long vowels used in this study, only two, /ø:/ and /ɑ:/, are in the centre of the F1/F2 vowel space. Three inter-category continua used in this study include one or more of these centrally located vowels. Two of those continua—/o:/ → /ɑ:/ and /ʉ̟:/ → /ø:/—are associated with a significant decrease in discrimination performance. The third contrast from this group, /ɑ:/ → /ɛ:/, was used as the reference in the logistic regression and is therefore associated with 0 change in discrimination performance. These findings are in line with (Polka & Bohn, 2003) who claim that vowels in the centre of the F1/F2 vowel space are less perceptually salient than peripheral vowels.

There are three continua in this study that only include vowels near the edge of the F1/F2 vowel space. Interestingly, the discrimination performance was not the same among these continua. The /y:/ → /ʉ̟:/ continuum increased the discrimination chances more than /u:/ → /o:/ or /ɛ:/ → /e:/. Typically, /i/, /u/, and /a/ enjoy the status of focal vowels because they occupy the outermost corners of the F1/F2 vowel space. However, Polka & Bohn (2011) point out that /y/ also approaches formant convergence limits and maximum perceptual salience, and suggest that /y/ could be treated as a focal vowel along with /i/, /u/, and /a/. They base their suggestion in part on their earlier observation that /y/-/u/ contrast remains discriminable for adult speakers of English, where this contrast is not phonemic (Polka & Bohn, 1996). The finding of the current study may be interpreted as a supporting evidence for this line of thought.

This study did not include the continuum /ʉ̟:/ →/u:/. This continuum has a long vector, connecting two salient, focalised vowels, and does not cross the central area of the F1/F2 vowel space. It would therefore be useful to include this continuum in future studies of this kind, to attempt to discern the effect of peripheral category location from the effect of inter-category vector length.

Dialect appears to be a significant factor in discrimination performance. Sydsvenska and Norrländska groups performed worse than other dialectal groups, whereas participants from Västmellansvenska and Östmellansvenska dialectal groups performed very similarly. Dalabergslagska dialectal area demonstrated better discrimination performance than any other dialectal area group. However, the results for the Norrländska and Dalabergslagska groups should be treated with reserve, because the number of participants in these groups was two and one, respectively. Such small group sizes do not allow to conclude whether the performance levels observed in the underrepresented groups are due to individual variation or reflect the true effect of those dialects. The remaining three dialectal groups can be discussed with more certainty. The similarity in performance of participants representing Västmellansvenska and Östmellansvenska suggests similar phonemic inventories of vowels in these areas. This finding would speak in favour of treating these regions as parts of the larger Central Swedish dialect family, at least when vowel perception is concerned. The Sydsvenska group discrimination performance was potentially affected by the diphthongisation which is prominent in greater part of the Sydsvenska geographical area. In natural speech, nearly all vowels are realised in a non-flat shape due to formant transitions to and from the surrounding segments, so it is possible that all participants were affected negatively by the flatness of the synthesised vowels in the stimuli. But in a large part of the Sydsvenska dialectal area Scanian dialect is spoken, where vowels commonly take on very dynamic shapes, often approaching diphthongs. Compared to the rest of the listeners, perception in participants from the Sydsvenska group may have been hampered for this reason.

We also observed that discrimination performance was better for stimuli around the categorical boundaries compared to stimuli at the edges of the continua, close to the prototype vowels. As such, this finding is not surprising, because categorical perception indeed implies suppression of acoustic discrimination acuity around the prototype vowels in native speakers. However, participants of the current study demonstrated performance variation which can be attributed in part to dialectal areas. If the dialectal areas had sufficiently dissimilar phonemic inventories, one might expect the boosting effect of categorical boundaries to dissolve. The fact that the boundary effect remained distinguishable can have different explanations. On the one hand, it could be a sign that the phonemic inventories of the Swedish dialects are not so dissimilar after all. Also, the stimuli of this study were based on East Central Swedish vowel categories, while two thirds of the participants of this study were raised with predominantly East or West Central Swedish dialectal influence (which have presently been associated with a very similar discrimination performance). Thus, only a third of the participants may have been contributing responses that had a potential to smooth away the effect of categorical boundaries. If that is the case, one would expect to see a greater effect size for categorical boundaries if the experiment only included Central Swedish participants. Another explanation for the presence of the boundary effect concerns the age of the participants and urbanisation level of the area where they were raised. A recent comparative analysis of recordings of rural Swedish dialects found that younger speakers (aged 30-45 now) showed less geographical variation in vowel pronunciation than older speakers (aged 65-85 now) (Leinonen, 2010). The results of the current study represent thirty young speakers aged 18-39, of whom eleven stated a major city (Stockholm, Göteborg, Malmö, Uppsala, Linköping) as their place of upbringing. Thus, the observed effect of the categorical boundaries in the present study may be an indication of the phonemic inventory consolidation process that is happening in modern Swedish dialects.

Conclusions

Both phonetic and phonological effects on discrimination of vowels in Swedish listeners have been found. Vowel category location in the periphery of F1/F2 vowel space was found to increase the discrimination performance. This was particularly the case with the front rounded vowels. Dialectal background also affects the discrimination function of native Swedish speakers. South Swedish dialectal area was associated with a decreased discrimination performance, which may be due to the diphthongisation that is prevalent in these dialects but was not present in the experiment stimuli. Continuous exposure to foreign languages (excluding English) does not appear to have a significant effect.

Acknowledgements

This work was supported by the Swedish Research Council (Grant No. 2018.00632), Knut and Alice Wallenberg Foundation (Grant No. 2018.0454), Crafoord

Foundation (Grant No. 2017.0006) and Marcus and Alice Wallenberg Foundation (Grant No. 2018.0021).

References

Bruce, G. (2010). *Vår fonetisk geografi*. Studentlitteratur.

Cabrelli, J., Luque, A., & Finestrat-Martínez, I. (2019). Influence of L2 English phonotactics in L1 Brazilian Portuguese illusory vowel perception. *Journal of Phonetics*, *73*, 55–69. doi: /10.1016/j.wocn.2018.10.006

Chistovich, L. A. (1985). Central auditory processing of peripheral vowel spectra. *Journal of the Acoustical Society of America*, *77*(3), 789–805. doi: /10.1121/1.392049

Escudero, P., & Williams, D. (2012). Native dialect influences second-language vowel perception: Peruvian versus Iberian Spanish learners of Dutch. *The Journal of the Acoustical Society of America*, *131*(5), EL406–EL412. doi: /10.1121/1.3701708

Gorba, C. (2019). Bidirectional influence on L1 Spanish and L2 English stop perception: The role of L2 experience. *The Journal of the Acoustical Society of America*, *145*(6), EL587–EL592. doi: /10.1121/1.5113808

Junge, C., Boll-Avetisyan, N., & Benders, T. (2019). Speech perception and discrimination: From sounds to words. In J. Horst & J. von Koss Torkildsen (Eds.), *International Handbook of Language Acquisition* (pp. 153–172). Routledge.

Leinonen, T. (2010). *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects*. University of Groningen.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. doi: /10.1037/h0044417

O'Brien, G. E., McCloy, D. R., Kubota, E. C., & Yeatman, J. D. (2018). Reading ability and phoneme categorization. *Scientific Reports*, *8*(1), 1–17. doi: /10.1038/s41598-018-34823-8

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. K., & Peirce, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195–203. doi: /doi: /10.3758/s13428-018-01193-y

Polka, L., & Bohn, O. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *The Journal of the Acoustical Society of America*, *100*(1), 577–592. doi: /10.1121/1.415884

Polka, L., & Bohn, O. S. (2003). Asymmetries in vowel perception. *Speech Communication*, *41*(1), 221–231. doi: /10.1016/S0167-6393(02)00105-X

Polka, L., & Bohn, O. S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*(4), 467–478. doi: /10.1016/j.wocn.2010.08.007

Schwartz, J.-L., Abry, C., Boë, L.-J., Ménard, L., & Vallée, N. (2005). Asymmetries in vowel perception , in the context of the Dispersion-Focalisation Theory. *Speech Communication*, *45*, 425–434.

Sun, Y., Lu, X., Ho, H. T., & Thompson, W. F. (2017). Pitch discrimination associated with phonological awareness: Evidence from congenital amusia. *Scientific Reports*, *7*(February), 1–11. doi: /10.1038/srep44285