

UNIVERSITEIT VAN AMSTERDAM  
Graduate School for Humanities  
MA Programme in General Linguistics

## **Portugal: Europe's Mistaken Identity**



Chris Deacon

Student Number: 11104015

[deacon17@hotmail.co.uk](mailto:deacon17@hotmail.co.uk)

Supervisors: Paul Boersma

Silke Hamann

Amsterdam

2016

Portugal: Europe's Mistaken Identity  
by Chris Deacon

### 1. Introduction

There seems to be a general consensus amongst those not well versed in Portuguese that it is a bit of an anomaly as a Romance language as it is perceptually not a Romance language at all. It is in fact perceived by many as "Eastern European". This is quite an assertion to make, especially with no well known academic sources to back it up. The following paper sets out to not only empirically validate this claim, but also investigate the reasons why. Through the use of an online questionnaire, this hypothesis will be examined together with three subsidiary hypotheses :

Main Hypothesis: Portuguese is perceived as Eastern European

Hypothesis One: Portuguese is perceived as Eastern European due to it being a stress-timed language.

Hypothesis Two: Portuguese is perceived as Eastern European due to its high proportion of sibilants.

Hypothesis Three: Portuguese is perceived as Eastern European due to its nasal vowels.

### 2. What is an "Eastern European" language?

Before the investigation can begin, it is necessary to define "Eastern Europe". Eastern Europe is traditionally thought to contain the following countries: Russia, Estonia, Lithuania, Latvia, Belarus, Ukraine, Moldova, Romania, Bulgaria, Albania, Poland, Hungary, Czechia, Slovakia, Slovenia, Croatia, Bosnia and Herzegovina, Montenegro, Serbia, Macedonia, Kosovo, as well as, arguably, Georgia, Armenia, and Azerbaijan.

What these countries all share in common is being on the east side of the "Iron Curtain". As Winston Churchill said in a telegramme to President Truman in May 1945: "An iron curtain is drawn down upon their front. We do not know what is going on behind." The expression stands as a "symbol for Russia's prevention of the free flow of ideas and information" (Feuerlicht 1955). In other words, these countries were, if not part of the Soviet Union, then to some extent controlled or influenced by it until its break-up in 1991.

21 years later the point is made in a blog featured on the Economist website entitled "The 'East' is dead: Time to scrap 'Eastern Europe'" (P.T.W. & E.L. 2012) that the concept of "Eastern Europe" is out of date. It contends that these countries were not all completely assimilated under the Iron Curtain; Eastern Europe contained "mavericks" like Romania, and "rebels" like Yugoslavia and Albania. Geographically this concept of Eastern Europe does not hold water either as Finland and Turkey (two of the most eastern countries in Europe) are not considered to be part of this block and Vienna is east of Prague. In addition, the blog claims that Eastern Europe is a "bad brand", as it carries connotations of "poverty, marginalisation, and weirdness", which is far from the truth as these countries have caught up and/or overtaken the rest of Europe in poverty indexes.

Perhaps then it is easiest to describe Eastern Europe linguistically. David Crystal (2010:p308) maps out the language families as follows: Portuguese belongs to the Romance or Italic family, as do Italian, French, Spanish, and Romanian; the Balto-Slavic languages consist of Czech, Polish, Latvian, Lithuanian, Bulgarian, Macedonian, Serbian, Croatian,

Slovene, Russian, and Ukrainian; and finally the Uralic languages (which aren't even Indo-European) consist of Estonian, Finnish, and Hungarian.

So the languages present in the countries of Eastern Europe are in fact drawn from all three language families. Clearly the majority of Eastern European languages come from the Balto-Slavic family but for the purposes of this study the "controversial", "outdated", and "politically incorrect" definition of Eastern Europe and its corresponding languages will be used. This is, after all, an investigation into language perceptions and the attitudes of the general public, not those of seasoned linguists.

The reasons why Romanian, Estonian and Hungarian but not Finnish will be included will be looked at in further detail in the methodology section.

### 3. Previous Language Attitude Studies

Before embarking upon this enquiry into attitudes about the Portuguese language, one must first consider previous studies. According to Rebecca Agheyisi and Joshua A. Fishman (1970), there are three categories into which language attitude studies are divided:

- "Those dealing with language-oriented or language-directed attitudes;
- Those dealing with community-wide stereotyped impressions toward particular languages or language varieties (and, in some cases, their speakers, functions, etc.);
- Those concerned with the implementation of different types of language attitudes"

Language attitude studies use a more sociolinguistic approach than the typological approach favoured by this study. "Impressions" and "implementations" seems to imply situations that have more to do with dialects and sociolects than languages. For example, nowadays people have more to say about what constitutes a superior dialect than a superior language. One could say that we have moved on from being xenophobic to being classist. So these attitude studies mostly look at judgements; the "impression" is evaluative rather than analytical. "Implementation" could refer to how an overt prestige is set as an educational template.

The typological approach that this study pursues demands from its participants analysis rather than evaluation, even if the analysis is a result of evaluative stereotypes which most attitude studies are attempting to discover. It is thus not a typical attitude study, and it is difficult to find a parallel.

Perhaps the closest approach Agheyisi and Fishman mention is the community-wide, stereotyped impressions toward particular languages. Clearly they are still taking a sociolinguistic approach, which is likely to dwell on concerns such as why people think that the languages of primitive people are less complex, or to take a more practical approach such as why some students lack the motivation to learn a particular language in a classroom environment.

The current study takes a step back from these typical studies' approaches, and takes a more general approach: why people have ideas or conceptions about what a Romance language should sound like or what a Balto-Slavic language should sound like. The current study's interpretation or adaptation of Agheyisi and Fishman's second category is seemingly not one they would have intended, as studies of the type they draw upon are, according to them, "generally concerned with the social significance of languages or language varieties". The difference in structure between typical attitude studies and the current one could be analysed algebraically. Taking a "typical study" as an example, Trudgill and Giles (1983:p219) found that British people considered rural accents of British English much more

pleasant than urban accents due to their “romanticised nostalgic view of the countryside”, whereas this effect was not found with Americans who do not know which accents are rural and which ones are urban. With both nationalities the structure of the experiment was “evaluate x given y.” The difference being that Americans only knew that y was English, whereas British listeners had more of an idea about which dialects were being spoken. In a sense the current study is investigating the naivety of Americans more than the prejudices of Britons. However, the structure is reversed; the participants are not being asked for x, they are being asked for y and they do not know what y is. In other words the experiment is “evaluating/analysing x, what is y?”.

Hypothesis one therefore, is that participants will implicitly analyse x as stress-timed and therefore decide y is “Eastern European” as the best fitting Indo-European language.

Back in 1928, sociologist Read Bain said “attitude is often used a synonym for habit”. This serves as a useful starting point. Agheyisi and Fishman also agree that attitude is a subjective bias built upon through experience: “practically everybody agrees that attitudes are learned from previous experience, and that they are not momentary but relatively enduring.” Because of this, it is empirically difficult to measure. Agheyisi and Fishman also highlight this point with two questions:

- “What constitute the right kind of data from which attitudes may be inferred?”
- How does one measure in physical terms (via observations, inquiries, or quantifiable measures) something which has no overt substance?”

The current study looks at differences between languages, which are more apparent, rather than the differences between dialects or sociolects, making Agheyisi and Fishman’s concerns slightly easier to tackle.

So there are two approaches one can take, and which have been taken previously in attitude studies. The first is to take attitude as a hypothetical construct; a subjective but independent variable. This approach has been criticised but the advantage of it is that this variable is a psychological constant not affected by the test conditions. The current study’s test conditions are essentially equal for all participants however, so the second approach, taking attitude as a dependent variable, is a better option. Taking attitude as a measureable, observable phenomenon gives the study an empirical robustness. Bain makes the point that “the only way to determine attitudes is by observation and statistical treatment of behaviour.” The criticism of this approach, that “attitude has no independence of the specific stimulus situations in which the responses are observed” (Alexander 1967), is not relevant to the current study, where all participants are tested in identical conditions.

One final point worth mentioning is that language attitude studies have been based on two theoretical approaches: the behaviourist approach and the mentalist approach. The former looks at overt behaviour; mainly in the responses people make to social situations. The latter looks at more covert behaviour, as attitudes are viewed as an internal mental state which may give rise to certain forms of behaviour. (Ihemere 2006).

It is the latter approach that the current study will take, as one’s perception of Portuguese as either Romance or Slavic reflects an internal mental state, while behavioural responses are not being looked at.

#### 4. Where did the hypotheses come from? Brief Introduction

There could, of course, be many other reasons why Portuguese is perceived as Eastern European, but the following hypotheses are an appropriate starting point for seemingly uncharted territory:

Main Hypothesis: Portuguese is perceived as Eastern European

Hypothesis One: Portuguese is perceived as Eastern European due to it being a stress-timed language.

Hypothesis Two: Portuguese is perceived as Eastern European due to its high proportion of sibilants.

Hypothesis Three: Portuguese is perceived as Eastern European due to its nasal vowels.

Before commencing with the methodology, it is necessary to justify where these hypotheses have come from, and briefly mention how it is possible to investigate them:

Subjects will be given six excerpts to listen to and they will have to decide which language they think it is. In investigating hypothesis one, European Portuguese will be used with Brazilian Portuguese as a control since the former is stress-timed and the latter is not. Hypothesis two's investigation will use European Portuguese with no sibilants and an Eastern European language heavy in sibilants as a control (Russian). And with hypothesis three, European Portuguese with no nasal vowels will be used together with an Eastern European language heavy in nasal vowels as a control (Polish). The unaltered original first excerpt will also be used as a control for the nasal vowel and sibilant free variants.

#### 5. Portuguese's Mutual Intelligibility with Spanish

It is necessary to justify these hypotheses, but there is one question that must first be considered: why doesn't Portuguese sound like its closest relative and Iberian Peninsula neighbour; Spanish? In fact it probably does, to Portuguese and Spaniards who claim to have some level of mutual intelligibility.

The British linguist William Entwistle (1953:p31) claims that what makes two speech forms languages rather than dialects is not mutual ease or difficulty of understanding, otherwise Spanish and Portuguese would be considered dialects of the same language (as would Norwegian and Swedish for example). Jepsen and Biel (1986) claim the understanding is unidirectional: "Most Portuguese have a fairly good, natural comprehension of spoken Spanish. But be forewarned that the reverse is not the case. Knowing Spanish will put you into a unique position for one-way communication- able to ask directions or make reservations but unable to understand the response." An experiment by Jensen (1989) discovered understanding or mutual intelligibility to be at around 50-60% although in this case the test conditions were not favourable for intelligibility (there was no face to face communication) but the test conditions were favourable in the sense that South American dialects of Spanish and Portuguese were tested rather than European dialects. The justification for hypothesis one will suggest why this might be the case. Jensen also agrees with Jepsen and Biel that Portuguese speakers understand Spanish more than the other way around.

Two specialists in Portuguese, Ellison and Andrews (1969:p259), make an interesting point about mutual intelligibility:

“Who is not familiar with the notion-often an abused one - that, for a person who knows Spanish well, one week's thumbing through a Portuguese grammar is usually enough for a mastery of the written language? ... This is anything but the case, of course, where phonology is concerned: here the two languages are not mutually intelligible; rather they are remarkably far apart.”

Spanish and Portuguese are very similar orthographically but phonologically they are worlds apart, or at least on the other side of the continent from each other. The two sentences below are translations of the English sentences “Spanish and Portuguese are generally quite different.” They look orthographically similar, but Portuguese has more vowel elision and a completely different pronunciation of <ge>

<Español y portugués son en general bastante diferente>  
/espa ol yi po tu es son en xene al βas tante ðife ente

<Espanhol e Português são geralmente bastante diferente>  
/ p li pu tu e s ãw alm êt b t ãt dif êt /

## 6. The History of the Iberian Peninsula

How did two languages, so close geographically and orthographically, diverge so much phonologically? To answer this question, it is necessary to consider the history of the Iberian Peninsula.

Portuguese and Spanish are of course both descendents of Latin after the Romans settled there in 218 BC and colonised it in two centuries. The exception being the Basque country, which has its own culture and language completely unique from any other. Iberia was divided

into four provinces: Tarraconensis, Baetica, Lusitania, and Gallaecia. (Azevedo 2005). The latter region was named after its early inhabitants, the Gallaeci. Over a couple centuries, the Latin spoken in Gallaecia changed into two languages: Galician, which is still spoken in that region, and Portuguese, which spread southwards into the kingdom of Lusitania, the prefix of which generally refers to Portuguese (ie “lusófono” meaning Portuguese speaking and “lusofonia” meaning Portuguese people or regions).

The Romans were not the only settlers. Many Germanic tribes also conquered the plain. Most notably, through the 5th to the 8th century, the land which constitutes modern Spain was part of the Visigothic Kingdom. The Visigoths were “displaced by the sword of Islam”. (King, 1972).

By the tenth century, Arab conquest of the Iberian Peninsula contributed to five small northern territories speaking different Romance varieties (as well as the Basque region). The three territories and language varieties which had the most success reconquering the south were Catalan, Castilian (which would extend over most of the peninsula and become modern Spanish) and Galician-Portuguese. Further expansionism by Afonso III, the fifth Portuguese king, led to consolidation of Portugal's borders and independence as well as the gradual differentiation between Portuguese and Galician. While Portuguese became the official language of an independent state, Galician was limited to oral communication at a local level. The northern Galician area would eventually become part of Castilian Spain. Systematic use of Portuguese rather than Latin also began under Afonso III's reign (1248-1279).

This historical background can help answer the question: why are Spanish and European Portuguese phonologically so different, but not so different orthographically?

It is probably because the printing press came into use in Portugal at around the same time as Portugal achieved independence from Spain. Once languages are in print and the spellings are distributed throughout the population, it is probable that the spelling does not alter a great deal, while Spanish and Portuguese pronunciation had nearly 700 years to diverge and shape themselves into their current form. European Portuguese in a sense became an oddity in that it didn't adhere to isogloss theory in the same way that Brazilian Portuguese did, which to the outsider sounds more like South American Spanish than European Portuguese. (This statement will be explored in the background to hypothesis one section). Portuguese colonisation of course gave birth to Brazil and Brazilian Portuguese after Portuguese settlement there in around 1500.

### 7. Early Sound Changes in Portuguese

The point was made that perhaps Portuguese broke away from Spanish after Portuguese began to be printed as an official language. Given that Brazilian Portuguese is more similar, for the outsider, to Spanish than European Portuguese is, one would think that this divergence happened mainly in the late history of European Portuguese. But perhaps not, perhaps Brazilian Portuguese was influenced by Spanish in South America, and perhaps many of the sound changes which differentiate Spanish and European Portuguese happened before Portugal became an independent country, or indeed separated from Galician. Rip Cohen (2010) gives a detailed account of the historical phonology of Galician-Portuguese. The most relevant sound changes relate to the three hypotheses. (Stress-timing, sibilants, and nasal vowels). Processes that lead to a language becoming more stress-timed are vowel reductions or deletions and consonant clustering. One example was the loss of final vowel /i/ or /e/ after coronal consonants:

/i/ or /e/ > Ø / /l /r/ /n/ /z/ /s/ \_\_#.

Some initial vowels also fortified into consonants such as these two sound changes:

/i/ > /j/ > /d / or / /

/u/ > /w/ > /v/ > /β/

On the rise of sibilants (hypothesis two), Cohen asserts that fricatives in general appeared more regularly in Galician-Portuguese than in Latin where they were rare in medial position. Spirantisation in Portuguese could have led to more sibilants. Cohen gives some examples of spirantisation in Galician-Portuguese:

/p/ > /b/ (eg capere > caber) - this isn't strictly spirantisation even though Cohen claims it to be; perhaps the voicing of /p/ makes it more prone to further increasing sonority or spirantisation /β/.

/b > /v/ (eg dēbēre > dever)

Perhaps the most concrete sound change is of hypothesis three as "The fall of intervocalic n gave GP a new class of nasalized vowels (ã, ě, ĩ, õ, ũ) and diphthongs (õe, ãe)":

VnV > (nV)V (eg mănēre > mãer).

Listing all the sound changes from Latin to Vulgar Latin to Galician-Portuguese to modern Portuguese is outwith the scope of this study, and it is important to note that although the aforementioned sound changes seem to be characteristic of modern Portuguese, many of them also occurred in Spanish, meaning that any presuppositions of the two languages diverging early in their history should be taken with a pinch of salt.

### 8. Sound Changes in Portuguese continued: From Latin to Modern Portuguese

The above account gives a small insight into how some of the divergence of Portuguese and Spanish after their evolution from Latin could have occurred. Eric Holt gives a more comprehensive overview in his 2016 paper, "From Latin to Portuguese". The first of many sound changes on the path to Portuguese (Spanish was still on this path at this time) was the evolution of Latin vowels. Latin had five corner vowels (the same as Spanish does now, at least in terms of monophthongs) (Ladefoged & Johnson 2010:p227) with length creating a phonological contrast. Slowly, articulatory differences took over and the length contrast was lost. As Holt puts it:

"This abandonment of weight (mono- vs bimoraic status) as a phonologically independent feature of vowels yields the vocalic inventory of Hispano-Romance, which persists in Galician Portuguese and Modern Portuguese /i, e, , a, o, , u/" (But not Spanish).

This then led to "profound consequences for the consonantal system", which led to "the emergence of various identifying characteristics of Galician Portuguese." As the vowels lost their moraic status, so did the consonants, which became more sonorous and in the process acquired the features [+voice], [+continuant], and [+sonorant]. The loss of consonant moraicity can be seen in the degemination of "cuppam," "cattus", and "siccum" to "copa", "gato", and "seca". [+voice] and [+cont] can be seen in "lupum" and "rotam" to "lobo" and "roda" and "pedem" and "regem" to "pé" and "rei".

As Holt puts it:

"In the end, the drive to eliminate consonantal moraicity in its entirety has reshaped the consonantal system, and the result of these leniting processes is a consonantal inventory that no longer shows an imbalance of mora-bearing segments."

This of course leads to a change in syllable structure, and one could point to this sound change as a reason why Portuguese is the only stress-timed language in the Romance family. This change in syllable structure led to a more varied consonant inventory.

The loss of moraic consonants meant the variety and number of CC structures was reduced, with there being a loss of geminates and consonant codas. These sound changes prompted new consonant varieties to emerge. In particular, palatal consonants. Arguably the most significant hypothesis to be left out of the current study is an examination of palatal consonants. European Portuguese has more palatal consonants than its language family members and geographical neighbours, and Russian is well known for the same. Perhaps that could be a reason why they sound similar.

The consonant lenition previously mentioned led to not just more nasal vowels, but also diphthongs, and nasal diphthongs, as vowels came into contact with each other when the consonant eroded away. Holt dates three sound changes which are evident in Portuguese today that arose during the Renaissance period: the heavy velarisation of syllable final /l/, the retraction and frication of trilled /r/ to uvular / ʀ /, and the raising of pre-tonic vowels - /a/ to / i /, /o/ to / u /, and /e/ to / i /.

One final Renaissance divergence he mentions is the result of contact with Germanic tribes (such as the previously mentioned Visigoths who spoke a form of Vulgar Latin) who did not really reach far west enough to affect Galician Portuguese, but who did affect the other Iberian languages.

It wasn't just phonology that changed, it was also morphology and therefore morphophonology. Galician Portuguese has many more contractions than the other



languages of the Iberian Peninsula, particularly prepositions with surrounding words, usually with articles, but also with demonstratives, pronouns, adverbs and other forms. Holt gives an example of the article-preposition contractions (orthographically) below:

	<b>a</b> “to”	<b>de</b> “from”	<b>em</b> “in”	<b>por</b> “for/through”
<b>o</b> “the” (masc, sg./pl.)	ao(s)	do	no(s)	pelo
<b>a</b> “the” (fem, sg./pl.)	à(s)	da(s)	na(s)	pela(s)

Such widespread morphological contractions helped contribute to Portuguese’s phonological divergence if nothing else (such as making it more stress-timed, as contractions are a major property of stress-timed languages).

Another morphophonological change that helped create the Spanish-Portuguese divergence, was the evolution of some tenses. There was more semantic bleaching of auxiliary verbs from Old Spanish to Modern Spanish than there was in the evolution of Portuguese, which kept about it an air of ambiguity. The overall phonological effect of this could have resulted in Portuguese ending up with a higher %C (which, as will be mentioned later on, leads to a language being more stress timed). The semantic bleaching of auxiliary forms leads to more grammaticalisation. Grammaticalisation can lead to consonant erosion as content words turn into grammatical words which turn into clines which turn into inflectional affixes (Hopper and Traugott 2003). Many inflectional affixes in Spanish, for example, are just vowels, and some even involve the deletion of consonants in the process of conjugating them (for example, in both Spanish and Portuguese “to see” is “ver”, Spanish conjugation deletes the consonant; “veo” and Portuguese replaces it; “vejo”).

With Portuguese, more auxiliary verbs mean less grammaticalisation and therefore consonant erosion. Holt gives an example of this lack of semantic bleaching or “medioclisis” or “mesóclise” with “weak object pronouns between the verbal forms ‘dar te lo ei’ (give it to you I will) and ‘dar te lo ia’ (give it to you I was going). Holt makes one final point about grammaticalisation in Spanish compared to Portuguese which could contribute to their differing phonologies and therefore foreign perceptions of them. Latin forms of demonstrative pronouns and articles bore stress, but in Spanish these became prosodically weak and grammaticalised (so more consonant erosion). This weakening did not happen, at least not to the same extent, in Galician-Portuguese.

In Holt’s mind, many of the sound changes, which separated Portuguese from its Romance family members, happened in the early Modern period or before, most notably the evolution of syllable structure. This evolution has given rise to hypothesis one.

### 9. Background; Hypothesis One: Portuguese is perceived as Eastern European due to it being a stress-timed language.

European Portuguese is unique in being the only Romance language which is stress-timed. This is arguably why when outsiders hear it they can’t identify it as having any similarity to Spanish, French, or Italian. Hypothesis one is empirically testable because Brazilian Portuguese belongs to the same rhythmic class as all the other Romance languages (syllable-timed). This is unquestionably the main difference between the two dialects and most likely the most significant cue in distinguishing them.

According to Sónia Frota and Marina Vigário (2001) “this distinction has been around in the literature at least since the early eighties but no clear support for it, experimental or other, has been given” (and they also agree that European Portuguese is the only Romance stress-timed language).

Frota and Vigário measure the isochrony of European and Brazilian Portuguese, as well as some other languages, and plot them on a scale. Isochrony is not directly measurable so they base it on Ramus et al.’s (1999) acoustic measurements of consonantal and vocalic intervals. Ramus et al. successfully correlated these measurements with traditional rhythmic types. These traditional rhythmic groupings are as follows (Pike 1945, Abercrombie 1967:p97):

- Syllable timing; languages in which the duration of every syllable is equal
- Stress-timing; languages in which the interval between each stressed syllable is equal
- Mora-timing; languages in which the duration of every mora is equal. (Morae are found in languages such as Japanese).

Phonetic studies however, have failed to confirm that these durations are strictly equal and instead it would be more accurate to say that the timing regularities are based on syllables, interstress intervals, and morae. In fact, Dauer (1983) shows that in stress-timed English and in syllable-timed Spanish, the duration of interstress intervals is proportional to the number of syllables in the interval, and so interstress interval durations are just as constant in stress-timed languages as they are in any other. Dauer thus proposes that rhythmic distinctions are the results of specific phonological and phonetic properties such as syllable structure variety and complexity, vowel reduction, and the correlates of stress.

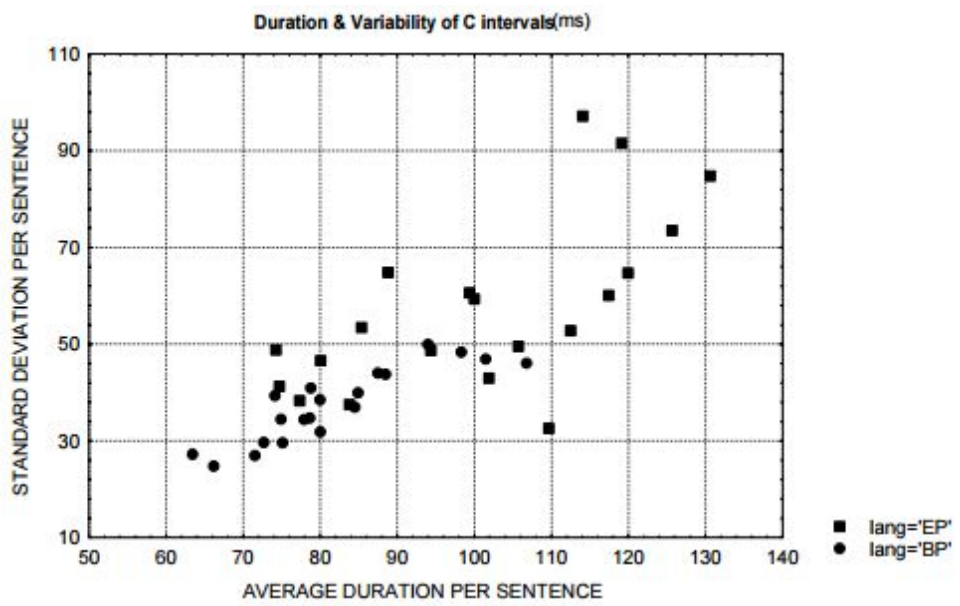
With regards to syllable structure, stress-timed languages tend to have much more syllable variety which is usually helped by the language allowing complex consonant clusters. The English word “strength” for example is one syllable with a vowel phoneme surrounded by five consonant phonemes. Syllable or mora-timed languages have more limited syllable types and are affected by phonological processes, which simplify syllable structure such as cluster simplification and epenthesis. Japanese, for example, is a mora-timed language and it only has about 50 different syllables due to the only syllable structure that is allowed being CV (“strength” is CCCVCC).

Regarding vowel reduction, stress-timed languages’ unstressed vowels tend to have a reduced vowel system and be phonetically shorter or absent, while in syllable-timed languages, unstressed vowels are usually not centralised nor are they consistently shorter. Finally, when it comes to the correlates of stress, stress-timed languages display a lengthening effect, and stressed syllables carry the weight of intonation. By contrast, syllable-timed languages have no such lengthening effect, and intonation and stress are usually distinct and independent.

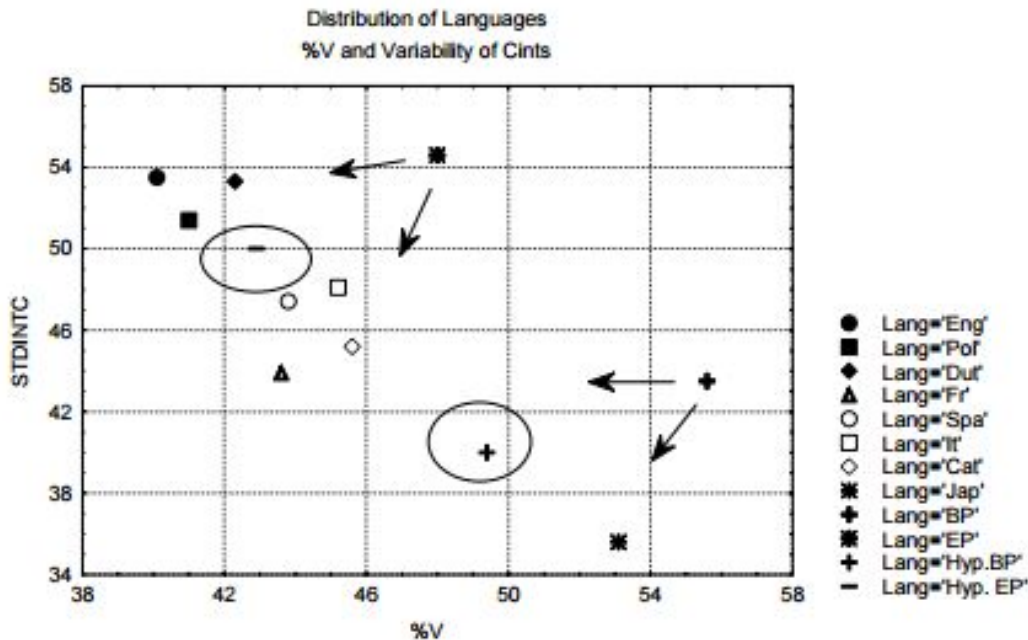
The combination of these properties make stressed syllables seem more stressed than unstressed syllables in stress-timed languages as compared to syllable-timed languages where all syllables seem equally salient. An example of the different phonological processes which distinguish European and Brazilian Portuguese is illustrated below (by Frota and Vigário). In European Portuguese, reduced vowels frequently delete yielding consonant clusters; on the other hand, in Brazilian Portuguese, vowel epenthesis splits consonant clusters and there is a tendency for coda loss:

European Portuguese		Brazilian Portuguese	
obedecer	mal	objetivo	mal
/obdse /	/mal/	/obi t ivu/	/maw/
“to obey”	“badly”	“goal”	“badly”

The measurements Ramus et al, and Frota and Vigário use for isochrony are the proportion of vocalic intervals within the sentence (%V), and the variability of vocalic and consonantal intervals expressed through a standard deviation measure ( $\Delta V$  and  $\Delta C$ ). Having said that,  $\Delta V$  is not a consistently accurate way to measure isochrony, as although reduced unstressed vowels characteristic of stress-timed languages help increase the variability of vocalic intervals,  $\Delta V$  is also influenced by other variables which do not relate to isochrony such as vowel lengthening processes and intrinsic vowel duration differences.  $\Delta C$  and %V correlate well with the rhythmic classes: stress-timing languages contain syllables containing more consonants and consonant intervals of greater duration (eg “strength”). Therefore they exhibit a greater consonant/vowel ratio expressed in a lower %V and a higher  $\Delta C$ . An example of the latter property empirically tested by Frota and Vigário is given below:



The  $\Delta C$  is measure on the y-axis and the average consonant interval duration per sentence is measured on the x-axis. It is quite clear that the stress-timed European Portuguese has more consonants per sentence and more complex consonant clusters. A more useful plot of  $\Delta C$  to %V is given below:



On this scale, English is deemed as the most stress-timed of all the languages having the lowest %V and the highest  $\Delta C$ , while Japanese is the least stress-timed language as a syllable-timed language is the middle ground with Japanese being a mora-timed language. In the graph, European and Brazilian Portuguese measurements were plotted by Frota and Vigário against a different experiment where different measurements were taken, so they added an estimate to place them more in line with where they would have been had the original measurements also measured Portuguese. Had Frota and Vigário also measured all the other languages on the graph, they might have aligned with Portuguese.

A problem with this approach to rhythm or isochrony can be seen in the graph: languages are scattered along a continuum rather than separated into the three distinct classes which this hypothesis is built upon. The above graph alone demonstrates this fact, with the estimate for European Portuguese falling halfway between the stress and syllable timed languages and also having a %V to  $\Delta C$  closer to the syllable timed languages than Brazilian Portuguese. However, several studies on language discrimination have shown that rhythmic distinctions play an important role in the perception of speech.

Frota and Vigário summarises these findings: "This line of research has shown that both adults and newborns, when exposed to filtered speech that preserves prosodic cues, are able to discriminate between languages belonging to different rhythm classes, but not between those of the same class. These findings strongly suggest that the properties behind the rhythmic distinctions are somehow encoded in the speech signal."

So there must be a kind of perceptual boundary that helps us distinguish between rhythmic classes (such as the differing voice onset time boundaries for Dutch and English with regards to aspirated /p/, unaspirated /p/ and voiced /b/). Additionally, because the perception of rhythmic classes is so strong that adults and infants can distinguish between them but not within them (Frota and Vigário), hypothesis one is indeed a very worthy hypothesis to be investigated.

### 10. Background Hypothesis Two: Sibilants

Hypothesis one's all encompassing nature means that hypotheses two and three are almost subsets of it. Hypothesis two of course is that Portuguese is perceived as Eastern European due to its high proportion of sibilants.

Luis M. T. Jesus and Christine H. Shadle (2002):

"Portuguese is unusually rich in instances of vowel reduction, consonant clusters, and plosives that are realized as fricatives."

Vowel reduction and consonant clusters relate to what constitutes stress-timed languages (as the spirantisation of plosives does if one considers the lengthening effect on the consonant interval). The latter is not really mentioned in isochrony literature so it is worth considering as part of hypothesis two because spirantisation leads to more fricatives, which of course means more sibilants. But where does the idea of more sibilants indicating a more Eastern European feel come from? Jaye Padgett and Marzena Zygis (2003) give the consonant inventories of Polish and Russian as follows:

#### Russian consonant phonemes

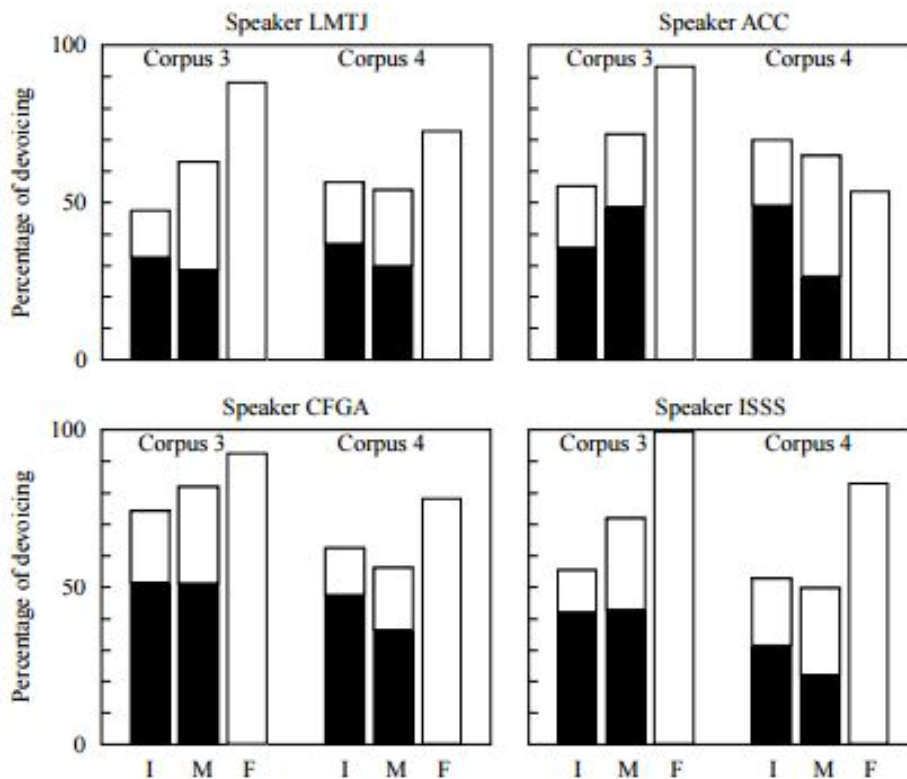
p p <sup>j</sup>	t t <sup>j</sup>	k k <sup>j</sup>
b b <sup>j</sup>	d d <sup>j</sup>	g g <sup>j</sup>
f f <sup>j</sup>	s s <sup>j</sup> (ç:) ş	x x <sup>j</sup>
v v <sup>j</sup>	z z <sup>j</sup> (z:) z <sub>l</sub>	
	ts t <sup>j</sup>	
m m <sup>j</sup>	n n <sup>j</sup>	
	l l <sup>j</sup>	
	r r <sup>j</sup>	
		j

#### Polish consonant phonemes

p p <sup>j</sup>	t t̄	k k <sup>j</sup>
b b <sup>j</sup>	d d̄z	g g <sup>j</sup>
f f <sup>j</sup>	s ç ş	x x <sup>j</sup>
v v <sup>j</sup>	z z̄ z <sub>l</sub>	
	ts t̄ç t̄ş	
	d̄z d̄z̄ d̄z <sub>l</sub>	
m m <sup>j</sup>	n ɲ	
	l	
	r	
		j
	w	

Russian and Polish, two of Eastern Europe's most widely heard languages in Western Europe, evidently have more sibilants than Romance languages, Polish has 8 or 10 sibilant phonemes and Russian has 12. How many does Spanish; Portuguese's closest relative have? Just two - /s/ and /t̄/. Portuguese may have only four consonant phonemes - /s/ /z/ / / / (Carvalho 2012:p20) but it has more in surface form and it is worth noting that European Spanish does not have palato-alveolar fricatives.

More importantly, it is the frequency of sibilants, not variety, that makes Portuguese resemble Polish and Russian more than Spanish. As previously mentioned, the writing systems of Portuguese and Spanish are very similar yet Spanish doesn't use as many sibilants. Orthographic {z} and {ç} {s} etc are all realised as sibilants. Another finding by Jesus and Shadle besides Portuguese's tendency to spirantise is Portuguese's tendency to devoice fricatives:



The overall results show that more than 50% of voiced fricatives devoice for all speakers except ACC. The graph above displays percentage of total devoicing by position in word; word-initial (I), word-medial (M) and word-final (F) of the fricatives /v/ /z/ and /ʃ/ combined. The black portion of each bar in the graph corresponds to fricatives in a stressed syllable, and the white portion to fricatives in an unstressed syllable. There are no Portuguese fricatives in final stressed position. This level of voicing, according to Jesus and Shadle, is very high compared with studies of other languages, and is an important characteristic of European Portuguese. Why is this relevant? As will be revealed, it is the voiceless sibilants that remind many people of Eastern European languages and in particular Russian. Secondly, this point ties into hypothesis one a little bit; devoicing of consonants reduces overall sonority in the acoustic signal, making it sound like there is a lower %V and thus more stress-timed.

As a final point, it is worth noting that the 18 second long European Portuguese excerpt used in the experiment for this study has 28 sibilants.

### **11. Background Hypothesis Three: Nasal Vowels**

Hypothesis three also ties into hypothesis one, as Portuguese does not just have more consonant (and in particular sibilant) variety than Spanish, it also has more vowel variety. However  $\Delta V$  turns out not be an accurate measurement for isochrony, and there is one particular aspect of the Portuguese vowel system that sets it apart from Spanish which fits the mould of this experiment and study: nasal vowels.

Spanish has quite a small vowel inventory, certainly in terms of monophthongs; it has the five corner vowels: /i/ /e/ /a/ /o/ /u/. Portuguese on the other hand has a slightly more complex vowel system (de Carvalho 2011):

## a. Monophthongs:

i            u  
 e            o  
 (ɐ)  
 ɛ            ɔ  
 a

## b. Diphthongs:

[iʊ]            uɪ  
 eɪ    eʊ            oɪ  
 (ɛɪ)  
 [ɛɪ] [ɛʊ]            [ɔɪ]  
 aɪ    aʊ

## c. Nasals:

ĩ                    ũ ũĩ  
 ẽ(ĩ)                    õ [õĩ]  
 ẽĩ ẽ õ

An example of an Eastern European language which it could have been confused with that has nasal vowels (and also central vowels) is Polish (Jassem 2003):

ĩ            ı            u  
 e ẽ                    o õ  
 a

Nasal vowels aren't restricted to Eastern European languages however; French is a good example of a Romance language that exhibits nasal vowels. It is worth noting that French is quite recognisable compared to Polish or Portuguese, particularly among the people who are tested in the experiment of this study (having English as a first or second language, French is also the second language learnt institutionally across English speaking countries such as the UK).

In addition - a point Jassem also makes - Polish nasal vowels have more in common with Portuguese than French. In Polish the phonetic realisation of a nasal vowel is often in actuality an oral vowel followed by a nasal semivowel. "Sa", for example, is pronounced [s ɥ̃] which is more similar to Portuguese "são" - [s ɥ̃] than to French "sont" [s ɔ̃].

Although this study is investigating why speakers of all languages - except perhaps Spanish and Portuguese - think Portuguese sounds Eastern European, it might be the case that people think Portuguese sounds Polish because of the similarity of nasal vowels.

At least half of the 200 participants taking part in the survey are British, and of all the Eastern European languages, they will have been exposed to Polish the most. According to BBC figures, Polish nationals represent the largest group of EU nationals living in the UK (853,000). This is a great deal more than the second largest group (331,000 Irish people) and the joint third group - Romania and Portugal, who have 175,000 each.

Hypothesis three is perhaps the least likely to be validated as the general public considers Eastern European languages to have lots of consonants, which sets Portuguese apart from its Romance neighbours, but the nasal vowel hypothesis is worth investigating as a control, if for no other reason.

## 12. Hypotheses 1, 2, & 3: Brazilian Portuguese

Although Brazilian Portuguese is only being used for hypothesis one, it is important to clarify its phonemic inventory in a way that relates to all three hypotheses:



EP consonants (Comrie 2009):

### PORTUGUESE

**Table 11.4 Portuguese Consonants**

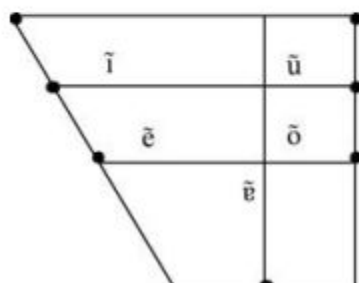
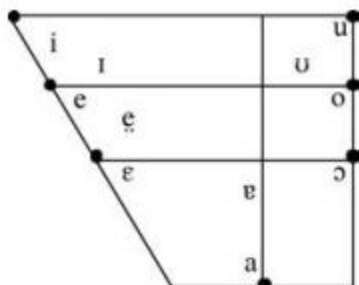
	<i>Bilabial (and Labio- dental)</i>	<i>Dental</i>	<i>Palatal (Palato- alveolar)</i>	<i>Velar</i>	<i>Uvular</i>
Plosives	p b	t d		k g	
Fricatives	f v	s z	ʃ ʒ		
Nasals	m	n	ɲ		
Laterals		l	ʎ		
Vibrants		r[r̄]			ʀ
Semi-vowels	(w)		(j)		

BP consonants (Barbosa 2004):

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Palatal	Velar
Plosive	p b		t d				k g
Affricates			(tʃ) (dʒ)				
Nasal	m		n			ɲ	
Tap				r			
Fricative		f v		s z	ʃ ʒ		χ
Lateral approximant				l		ʎ	

Coming back to H1, the very similar consonant inventories help point to the fact that the main differences between the dialects is stress-timing. So there are no real other factors at play. The main differences, it would seem, are that EP has a uvular trill, and BP has a velar fricative, and BP has dental affricates which EP lacks, so it could also work well as a control for hypothesis two.

BP vowels (Barbosa 2004):



The vowel chart to the left is the Brazilian Portuguese vowel inventory, and it is very similar to European Portuguese. Although not identical, they are the most closely related. The main difference being that EP has a high central vowel / ɨ / that BP lacks, a phoneme that is found in many Eastern European languages, so it could make EP sound more Eastern European than BP, but the effect would probably be minimal. Overall, this comparison adds further support to hypothesis one by showing that the main difference between EP and BP is stress-timing, and it also supports H3 in a more general sense in that Portuguese is perceived as Eastern European due to nasal vowels because their nasal vowel inventories are so similar too.



### 13. Methodology: The Questionnaire

The current study will model itself on the majority of language attitude studies and use an open question survey. Agheyisi and Fishman (1970) list the advantages of an open survey as opposed to a multiple choice:

- It encourages the subject to freely express their views without the surveyor prompting any answers.
- Subjects may reveal other attitudes which the researcher may not anticipate.

And the disadvantages:

- In a questionnaire, the effort needed to write down answers may make the respondent refrain from answering questions fully.
- Open ended questions can pose scoring problems for the researcher.

Using a questionnaire allowed a larger sample size, which would not have been possible with individual interviews. Secondly, although the questionnaire asks open questions, there are a limited range of possible answers, which allows quantitative analysis; yet participants were asked to justify their choices, giving qualitative data as well. Walter and Pronzato (1990), while comparing the merits of qualitative vs quantitative models, make the following point:

“At the risk of being slightly provocative, one might say that most experimenters think of experiment design as an art owing much to intuition and very little to mathematics, while mathematicians view it as an essentially statistical problem.”

It is fair to say that the current study’s methodological approach was more intuitive. However, it does satisfy the mathematician’s needs by having both quantitative and qualitative data. Quantitative in statistically analysing the responses of 200 people, nearly all of whom gave definitive answers. And qualitative in that many participants also justified their personal decision with an explanation, each of which is analysed.

### 14. Methodology: The Procedure

200 participants were recruited to take part in a “guess the language” survey. They were first asked for their name, age, first language, plus any other languages spoken and to what level. At the top of the next page, they were given an instruction: “You will be given 6 excerpts to listen to, you have to decide what language they are and why you think so. For each excerpt, press the play icon.” There were then 6 wav files uploaded on the page with the open question: “What language do you think this is? What makes you think it is that language?” written under each one with a paragraph sized textbox.

All excerpts was spoken by people aged 18-25. The first excerpt was 18 seconds long and was a recording of a young adult female from Lisbon, who read an extract from a random Portuguese Wikipedia page. The speech as a whole doesn’t make sense entirely as some of the recording was edited away to shorten the speech and to remove misleading proper nouns. The Portuguese and English transcript is as follows:

“As tropas israelenses invadiram o sul do Líbano em 14 de março de 1978 e chegaram até o rio Litani, com o objetivo de liquidar as bases da Organização de Libertação da Palestina no país. A partir do sul libanês, a guerrilha palestina destinadas pela O.N.U. a um futuro Estado palestino”

Translation: “The Israeli troops invaded the south of Lebanon on the 14th of March 1978 and reached the Litany river with the goal of liquidating the bases of the Palestine Liberation Organisation. From the South, the Palestinian guerrillas designed by U.N. to become a future Palestinian state.”

The second excerpt was the only one which was not recorded personally by me on Praat. It is a Youtube video of the Brazilian professional footballer Neymar converted into a 12 second Wav file. Neymar is said to speak a working class São Paulo dialect. The transcript and translation is as follows:

“Que chegue logo a hora do jogo, né! Mas vamos trabalhar aí, treinar bastante, para que esteja preparado para o dia. A torcida comparecendo no estádio, nos ajudando e gritando, o time vai pra frente e vai ganhar esse campeonato.”

Translation: “May the match time come soon, right! But, let's work hard, train a lot to be prepared for the match day. With the fans showing up at the stadium, helping us and screaming, the team will move forward and will win this tournament.”

The third excerpt involved European Portuguese with no sibilants. The transcript was formed through copying and pasting parts of sentences with no sibilants from random Portuguese Wikipedia pages. As such, the recording makes very little sense. The recording was made by a female from Lisbon and is 18 seconds long:

“Também é com uma crina de cavalo para confundir em que a partícula é o fotão. Não há um único irmã também foram em vinil como. Com o então bi-campeão mundial À nona prova da temporada, durante a corrida eu vou comprar um novo pau pão. Eu leio, te e bem-vinda.”

Translation: “It is also with a horsehair to confuse where the particle is the photon . There is no single sister were also in vinyl like. With the then world champion the ninth race of the season, during the race I'll buy a new stick bread. I read, you and welcome.”

The fourth excerpt involved no reading and was the 13 second recording of a Russian woman:

“Мне 18 лет. В этом году я начала учиться, но мне, как минимум, учиться тут 4 года. Мне, в принципе, тут достаточно нравится. И думаю, сейчас я перейду к части, где надо читать, потому что я понятия не имею, о чём говорить.”

Translation: “I'm 18. This year I have started my studies, and it will take me at least 4 years. But I like it here. And now I will switch to the reading part, since I have no idea what to talk about.”

The fifth excerpt was similarly constructed to the third one; parts of a Wikipedia page patched together, this time free of nasal vowels. It was also possible to form a very basic introductory sentence (17 seconds long in total):

“Eu gosto de correr ao redor da cidade, e de nadar e ler texto aleatório. Eu quero que você me ajudar a falar, aqui está.

A ideia repetida pode ser rítmico o corredor leste-oeste. É plausível que as propriedades biológicas fica disposto sobre a roda prata sobre o arcebispo e eleitor.”

Translation: “I like to run around the city, and to swim and read random text. I want you to help me to speak here .

Repeated rhythmic idea can be the east-west corridor . It is plausible that the biological properties is arranged on the silver wheel on the archbishop and elector.”

The final excerpt was a Polish man reading a book blurb (16 seconds long):

“Opis książki: Hit dla wielbicielek Imperium Osmańskiego i Sulejmana Wspaniałego. Książka, na podstawie której powstał serial produkcji tureckiej - Wspaniałe Stulecie. Sułtanka Kosem, następcza tronu to niezwykła postać”

Translation: “Book Description: A must-read for those who are interested in the Ottoman Empire and Suleiman the Magnificent. The Turkish Production TV series, The Great Century, was based on this book. Empress Kosem, the next throne successor is an unusual person.”

## **15. Demographics**

Once 200 people had completed the survey, it was important to make sure that it had covered a large demographic, so that any conclusions drawn could be extrapolated to a wider population.

There was a good spread across both genders with 108 females, 89 males, and 3 subjects who did not specify. Although the majority of participants came from a younger demographic, there was still a decent proportion of older participants, with 95 subjects aged 18-25, 50 aged 25-40, 20 aged 40-55, 20 aged 55 or older, and 15 participants who did not list their date of birth.

Perhaps the most important demographic was first and second languages spoken. Judging fairly subjectively from people’s responses in their descriptions of their second language ability, a 35-40% estimate of the total group being monolingual is a fair assumption to make. It is also safe to assume that everyone who took the test spoke English to at least a conversational level.

137 participants spoke English as a first language, 30 Dutch, 5 Italians, Germans, and French each. There were also 4 Poles, 3 Greeks, 3 Russians, 3 Romanians, 2 Bulgarians, 2 Portuguese, and 2 Spaniards. There was one speaker of each of the following languages (as a first language): Afrikaans, Bosnian, Czech, Gaelic, Norwegian, Slovene, Swedish, Vietnamese, and Arabic.

It is arguably a diverse enough demographic, linguistically, that one could draw conclusions about all Europeans’ perceptions of Portuguese (excluding of course Spanish, and Portuguese).

It is also important to note that 45 participants claimed to have some level of Spanish as a second language, and 5 claimed to have Portuguese, as well as 100 claiming to speak some

French, and 50 speaking some German. The diversity doesn't end there with the following also listed as second languages: Italian, Dutch, Bislama, Cantonese, Mandarin, Polish, Croatian, Serbian, Turkish, Russian, Swedish, Danish, Norwegian, Basque, Bahasa, Japanese, Thai, Catalan, Slovene, Latin, Gaelic, Gaidhlig, Romanian, Arabic, Swahili, Slovak, Korean, Pashto, and Urdu.

#### **16. Methodology: "Eastern Europe's" Boundaries and its empirical justification**

Before the results were analysed statistically, it was important to draw the boundaries of "Eastern Europe". As previously mentioned, the Balto-Slavic languages were an obvious candidate for this category, but why include the Romance language Romanian, and the Uralic languages Hungarian and Estonian, but not Finnish? This study is looking at why people think Portuguese sounds "Eastern European" so it is important to go with their definition of what constitutes an Eastern European language. Luckily, there is some data justifying their selections. Romanian in the end was included in the statistics as both a Romance language and an Eastern European language. Throughout the whole survey, Romanian was guessed 38 times and around a quarter of participants commented on its flexibility to fit both categories:

- "Romainian. [sic] Sounds Eastern European but not strong enough for Russia"
- "Romanian. Seems to have recognizable Latin roots (catorzh, paid) but Slavic intonation. Purely a guess. Maybe it's Catalan or something Portuguese derived."
- "Romanian; sounds a mix between a latin language with some eastern european sounds"
- "Rumanian Slavic and French sounds"
- "Romainean, [sic] has some North Eastern Europe influences but has the fluidity of a Latin based language"
- "Romanian. Sounds romantic but also sounds balkan."
- "Romanian as think might have Italian overtones but also sounded at times but Welsh"
- "Romanian. It sounds vaguely Latin a bit like Portuguese or Italian but then at times as an eastern sounding twang a bit like polish or Russian."
- "This sounds like something in between slavic and french so I'm gonna guess Romanian....."
- "Romanian? It is a romance language but it is def not italian and the intonation is strange."
- "Romanian. I have no idea what language this is, but it sounds a bit like italian."
- "Romanian? Bit Italian sounding but don't think it's Italian. Think Romanian [sic] has some similarities."

It is important to note that in total, around one half to two thirds of all participants made further comments justifying their selection, as opposed to just guessing the language and not writing anything else. The above evidence is clearly enough to allude to Romanian being both Romance and Eastern European; thankfully it is the only language which crosses into both language families.

In the case of the Uralic languages, Hungarian was unanimously considered to be an Eastern European language. Hardly surprising when you consider it is landlocked by 6 Balto-Slavic speaking countries and Austria and Romania. Hungarian was selected 25 times across the survey for the following reasons:

“East Europe? Hungary? Didn't recognise it but the pronunciation makes me think of eastern Europe.”

“Hungarian. Sounds like an eastern euro accent but sounds also very different from the others”

“Hungarian. No real idea, some of the words sounded nordic (kvall, noen, prøvde) but other bits sounded more romantic.”

“I don't know! This is really difficult, I reckon it sounds Eastern European. could it be hungarian?!”

“Hungarian. Sounds broadly Eastern European or Russian. Words ending in tsa.”

“Bulgarian or Hungarian, it sounds eastern european, but not quite Arabic”

“Hungarian. I don't recognize a word, so I guess something Slavic. But sounds more friendly and less 'sharp' than the last two”

“Hungarian. Eastern European”

“Quick fire syllables suggest Spanish or Italian, the 'eh...' While he is thinking suggests French. Also a hint of Russian. Hungarian? I've heard their language has links with Finnish. No, Hungarian would be too far East. Montenegrin?”

The last point is interesting, although it selects Montenegrin, not Hungarian, it alludes to the fact that Hungarian and Finnish are related. So why hasn't Finnish been included as an Eastern European language? For the following reasons:

“Finnish - a Nordic language I've never heard. The 'l' was very thin, the au and eu sounded Nordic. The musicality (higher and lower notes) of the words, less rhythmical. But it also sounded a bit southern european too.”

“Finnish, gut feeling.”

“Finnish, sounded Scandiwegian [sic]”

“Finnish, sounds Scandinavian but not Danish or I think Norwegian.”

“Finnish, sounds like a nordish language”

“Finnish. It sounds Nordic since there are a lot of really strange vowel sounds.”

“Finnish? Could be Scandinavian language?”

Throughout the survey, Finnish was selected 12 times and nearly every time it was justified as a Scandinavian language. Most people assume that due to Finland being part of Scandinavia, it has more in common with the North Germanic languages than the Uralic ones or even Russian, with which it apparently holds some level of mutual intelligibility. Lastly we come to the justification for categorising Estonia as Eastern European, the significance of which won't be very high as it was only guessed 7 times; four of which were justified as follows:

“Estonia. Or something eastern European”

“Estonian - because it also sounds russian”

“Estonian. Sounds a touch Nordic, but not Nordic. A guess based on Estonia's location.”

“Possibly Estonian? Reminds me of Hungarian a little bit but it also sounds like a Nordic language.”

Although two people suggested it could be Nordic like Finnish, the Eastern European categorisation is just about justified, for the sake of consistency, as otherwise it would be the only former Soviet State member not included in the “Eastern Europe” category.

### 17. Methodology of Results Section

The results were analysed in four sections:

- The first section: which simply analysed the main overall hypothesis as to whether EP sounded Eastern European or not.
- The second section looked at hypothesis one, analysing the results of excerpts one and two (EP and BP), the second of which was a control.
- The third section looked at hypothesis two, analysing the results of H2, using excerpt three (EP no sibilants) and excerpts one and four (Russian) as a control.
- The final section looked at hypothesis three, analysing the results of excerpt five (EP no nasals) using both Polish (excerpt six) and EP (excerpt one) as a control.

Main Hypothesis: Portuguese is perceived as Eastern European.

Hypothesis One: Portuguese is perceived as Eastern European due to it being a stress-timed language.

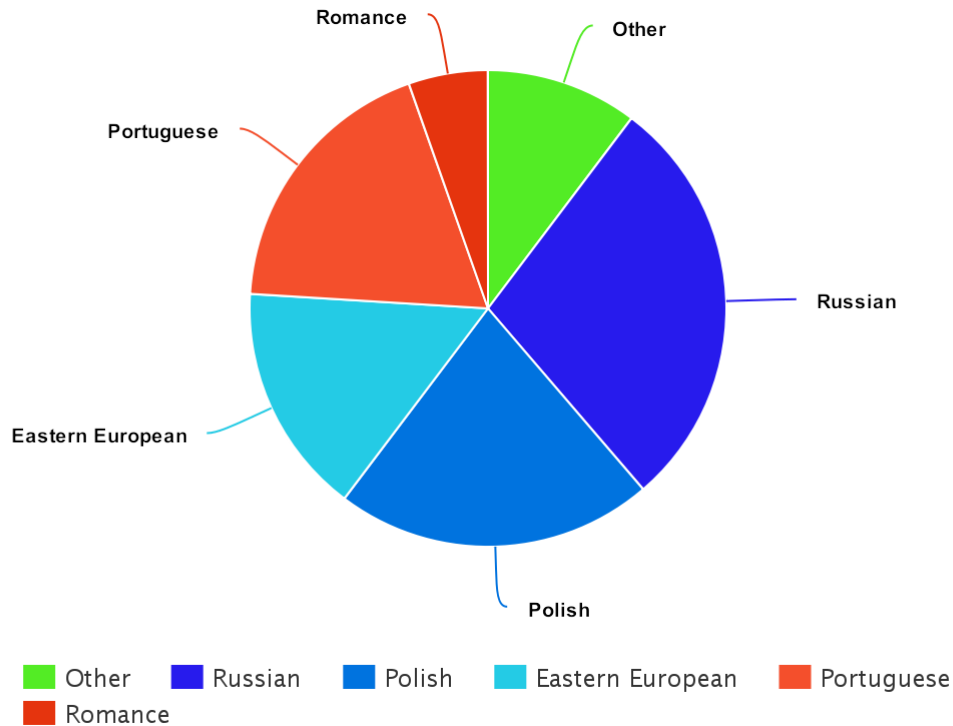
Hypothesis Two: Portuguese is perceived as Eastern European due to its high proportion of sibilants.

Hypothesis Three: Portuguese is perceived as Eastern European due to its nasal vowels.

### 18. Results: Main Hypothesis: Does European Portuguese sound Eastern European

Categorising results wasn't entirely straightforward. 200 people were tested but more than 200 languages were guessed for each excerpt. For example, one response was “I wouldn't know. Bosnian? Or maybe Serbian?” which would appear to constitute one guess for each of Bosnian and Serbian, but as both are Eastern European, it only adds one unit to the overall data. One slight inconsistency was that a Romanian unit was added to both Romance and Eastern European and then the number of Romanian units was subtracted from the overall total down to 200. (Since 10 people guessing Romanian would result in 10 added to Romance and 10 added to Eastern European, meaning a subtraction of ten from the overall number meant there were still 200 participants rather than 210).

This would slightly dilute the results of the guesses that weren't guessed to be Romance or Eastern European languages but Romanian was never guessed more than 10 times and the guesses for other languages weren't seriously analysed.



meta-chart.com

Above is the result of the first excerpt - unaltered European Portuguese. 137 out of 200 people guessed it to be an Eastern European language, 58 of those guessed Russian and 44 of those guessed Polish. 38 correctly guessed Portuguese. With over two thirds of participants believing Portuguese to be an Eastern European language this is strong support for the main hypothesis. A separate analysis was undertaken for each of the excerpts of the 73 participants who didn't have English as a first language. 38 of them thought the first excerpt was Eastern European, 22 thought it was Portuguese and 13 guessed other languages or didn't guess at all. That is just over a half of them who thought Portuguese was Eastern European rather than just over two thirds and more than half of the correct guesses (22 out of 38) came from the group of 73 (just over a third). This means that 99 out of 137 English as a first language speakers thought Portuguese was Eastern European, and only 16 of them realised or guessed it was Portuguese.

Around 60% of these English speakers were monolingual. As one would expect, there is an effect of knowing only one language, meaning they are more likely to perceive Portuguese as Eastern European. The overall results, and the results of the 73 in particular, would have made people in general and speakers of a first language other than English guess Portuguese correctly for a couple of extra reasons: The survey was open to everyone so a few people - perhaps as many as 10 - recognised the voices of one or two of the Portuguese girls being recorded, and over 50 participants claimed to have some knowledge of either Spanish or Portuguese (7 claimed to have at least basic Portuguese). There was no artificial filtering out of results of participants who knew the participants, Portuguese, or Spanish, but if there was, then the overall results might be even more spectacularly in favour of the theory that European Portuguese is perceived as an Eastern European language. Below are some individual responses to excerpt one which support this theory, without mentioning specific reasons why, which support the three hypotheses. A couple of them directly mention the fact that Portuguese sounds Eastern European:

“Either Russian or Portuguese. Has that sound about it .. I have heard Portuguese and it sounds Russian”

“It sounds Slavic to me, but it could very well be Portuguese.”

Many more of them support the premise of an “Eastern European” language family group, and most of these select Russian for that reason:

“Russian - it sounds eastern European”

“Russian as the accent sounds Russian or of a country near russia”

“Russian. It's got that slavic slur sound”

“Russian or one of the soviet nations. the accent.”

“Russian - actually I don't think it is Russian bit from the Eastern European area, I think it's that because there it has a European sound to it but also a lot of velar (back of the oral cavity) sounds to it which sounds a bit Russian”

“It sounds like a language from the east of Europe. The way the words are pronounced makes me think that. Im guessing Russian.”

“Russia. It sounds Slavic”

“Russian - Eastern european sounding”

“I recognise some broadly slavic sounds (vowels and endings of words) -- but difficult to specify. Russian?”

“Russian. Sounds slavic but I couldn't guess more accurately than that.”

russian: sounds like an eastern language

“Either Ukraine, polish, Russian. pronunciation, tone of voice.”

Many more support the premise of an “Eastern European” language family group, by guessing Polish, and placing it in the same group, even generalising to the point of comparing Polish to Russian, for example “Polish. Its sounds a bit like russian, eastern europe.” Comparing Polish and Russian, which are about as far as possible from each other in Eastern Europe - certainly Polish has the most geographically western Eastern European language, apart from perhaps Slovenian, which is rarely mentioned and the Sorbian dialects of Germany which are never mentioned; in terms of isogloss theory, the fact that these two languages are compared so apathetically solidifies the linguistic border created by this study. The evidence of this is below:

“Polish Sounds eastern European but not fast enough to be Russian etc”

“Polish, Eastern European accent but not strong enough to be russian”

“Polish. I only recognised 1 word and the accent sounded Eastern European, but not Russian. I'm fairly sure the word was Poland or Polish”

“Polish? It sounds slavic, but not Russian.”

“Polish or Russian - it just sounds like that - guttural bits”

“Polish sounds like a mixture of Spanish and German in a Russian accent.”

“Polish Eastern European harsh sounds”

“Polish. Sounds like an Eastern European language”

“I'd go for polish but I'm not certain. It's definitely 'eastern' if I can say so”

“Polish because some words sounded Eastern European”



“Polish? It sounds eastern-European. I have Polish friends, it sounds similar.”

“I think it's Polish, for I know it isn't Russian. Polish people tend to use their back part of the tongue more, I think (?)”

“Polish. At first I thought Russian, and then I heard slightly Germanic terms, for example 'stadt'. Felt as Poland was, historically at least, in the middle of the two it might be that.”

“Polish - sounds between russian and german”

Polish and Russian weren't the only Eastern European languages guessed, and more evidence for this family being consentient are given below, as well as more individual evidence for people thinking European Portuguese sounds Eastern European, but hard to place nonetheless (again because of its almost perceived homogenous sound to Western Europeans - some didn't even guess which language it was, eg “It sounds like a slave language”) - presumably that one meant to say slavic... it is unlikely the participant was guessing a creole!

“Czech, I think this because it has Germanic sounding pronunciation with certain romance influences, however it could just as easily be a Slavic language with romance influence.”

“Czech. Spunds [sic] Eastern European, I've been to Prague, sounds similar.”

“East Europe? Hungary? Didn't recognise it but the pronunciation makes me think of eastern Europe.”

“Slovakian? Not harsh enough for Russian but sounds like its from a similar region”

“Slovenian. Just a guess, sounds slavic and southern somehow”

“Slavian language. Serbian?”

“Serbian/Croatian/Bosnian. It sounds like a Slavic language to me because of the sound of the 'l' but to me it does not really sound Russian, it has something that makes me think it must be spoken further to the south.”

“Georgia - Because it sounds somewhat slavic in origin, but it's not Russian.”

“Slavic language. Not Russian, more balcanic. Because of the sound of the hard "L"”

“Albanian, or another eastern-europe language. I would say so because of the accent Latvian.....not really sure...sounds a bit like Albanian....which is quite soft in pronunciation and a bit eastern Or Central European sounding but can't place it.”

There were also a lot of vague justifications for which language was picked, such as “hearing it before” or it sounding “harsh”. Arguably one could argue that it sounding “harsh” supports hypothesis one as a “harsh” language is conceivably one heavy on consonants.

Nonetheless, the following individual responses give further insight to other cues, not part of the main or other three hypotheses that caused people to consider Portuguese to sound Eastern European:

“Russian - just sounds like it”

“Russian. The accent.”

“Russian. The phrasing and sounds.”

“Russian, harsh sounds”

“Russian - thick 'L' sound”

“Russian, the accent and pronunciation”

“Russian- it sounds how I imagine Russian would sound”

“Russian or Lebanse [sic]. It sounds Russian, but I thought I heard her say something with Lebanese.”

“Russian, it just sound like how Russian people talk”

“Russian, the pronunciation of the words and the sounds made.”

“Russian, accent and words sound russian”

“Russian, the accent sounds Russian.”

“Russian or Ukranian because I have a friend who speaks Russian and it sounds similar to that.”

“Polish -? Accent and intonation”

“Polish because it sounded similar to my uncle who was polish”

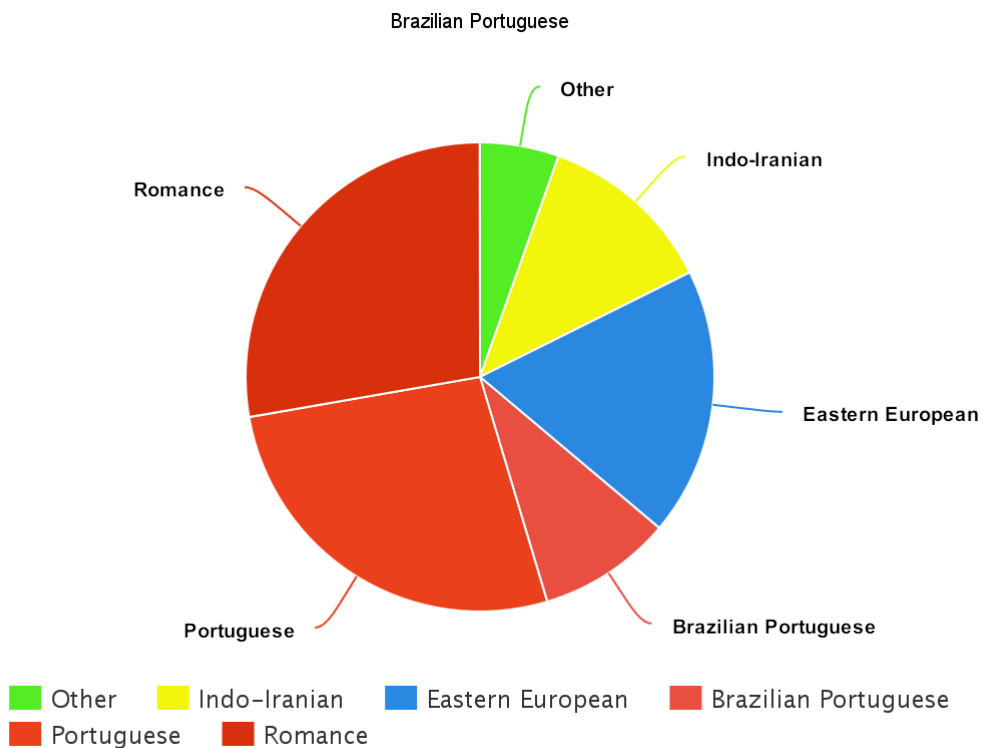
“Polish? Have heard Polish spoken and it had that ring to it”

“Polish - people who I have heard that are polish”

“Polish, they all sound alike.”

“Sounds Slovakian but cannot understand any words”

### 19. Results: Hypothesis One: Does European Portuguese sound Eastern European because it is stress-timed?



meta-chart.com

There is undoubtedly a huge difference between the perceptions of European Portuguese and Brazilian Portuguese. Only 26 people thought Brazilian Portuguese sounded Eastern European (as many as French). To put that into perspective, 137 people thought European Portuguese sounded Eastern European and only 3 people thought it sounded French, and the only Brazilian tested thought the European dialect sounded Russian: “Portugese (Portugal). I understand most of what she says. But honestly it could be two languages as I believe I should have understood more. It sounded like Portugese and Russian but it could be just Portugese. I'm from Brazil so the language there is a lot different.” The table below clarifies the figures:

	European Portuguese	Brazilian Portuguese
Portuguese	38	86
Eastern European	137	26

22 of the 86 correctly guessed that the dialect of Portuguese being spoken was Brazilian. The category Indo-Iranian was also included on the pie chart as 29 people thought the language belonged to that family (three more than Eastern European). Clearly there is a massive difference in perceptions of the two dialects, but an odds ratio test was carried out anyway to determine the statistical relevance of these results using the epitools package in R. (In the matrix, the figures 63 and 174 refer to the differences between 200 and number of people who picked an Eastern European language, or in other words, the number of people who didn't pick an Eastern European language):

```
$dataEPvsBP
```

```
> library(epitools)
```

```
> Data <- matrix(c(137,26,63,174),2,2)
```

```
> oddsratio(Data)
```

Outcome (original result of epitools package ran through R used diseases as coefficients, the edited table of the results are in the table below)

```
Predictor Disease1 Disease2 Total
```

```
Exposed1 137 63 200
```

```
Exposed2 26 174 200
```

```
Total 163 237 400
```

Outcome

	Eastern European	Not Eastern European	Total
European Portuguese	137	63	200
Brazilian Portuguese	26	174	200
Total	163	237	400

```
$measure
```

```
odds ratio with 95% C.I.
```

```
Predictor estimate lower upper
```

```
Exposed1 1.00000 NA NA
```

```
Exposed2 14.37331 8.758863 24.36533
```

```
Odds ratios with 95% confidence interval
```

	Estimate	Lower	Upper
European Portuguese	1.00000	NA	NA
Brazilian Portuguese	14.37331	8.758863	24.36533

```
$p.value
```

```
two-sided
```

```
Predictor midp.exact fisher.exact chi.square
Exposed1 NA NA NA
Exposed2 0 4.831346e-31 1.389549e-29
```

\$correction

[1] FALSE

attr(,"method")

[1] "median-unbiased estimate & mid-p exact CI"

	P value		
	Midp.exact (exact P value)	fisher.exact	chi.square
European Portuguese	NA	NA	NA
Brazilian Portuguese	0	4.831346e <sup>-31</sup>	1.389549e <sup>-29</sup>

The p-value of 0 (rounded) illustrates that there is a miniscule probability of these results occurring by chance. In addition, the high slope figure of 14.37 shows strong discrimination between the presence and absence of stress-timed syllables, lending strong support to hypothesis one if indeed the findings by many scholars on rhythmic classes are right that “both adults and newborns, when exposed to filtered speech that preserves prosodic cues, are able to discriminate between languages belonging to different rhythm classes, but not between those of the same class.” (Frota and Vigário).

Of course, one can argue that prosody is not the only difference between European and Brazilian Portuguese, they have many differences. However, the idea of what constitutes a rhythm class is so all-encompassing: it incorporates so many features of a language... syllable structure, consonant groupings, vowel reduction, or fortification, that such highly significant results must be down to the cue of syllable or stress (or mora) timing. If that is not enough evidence, then there are plenty of individual responses which comment explicitly on the way European Portuguese sets itself apart from its Romance family members by having the key feature of being a stress-timed language: a higher  $\Delta C$  and a lower %V:

“Russian: The same as with Ukrainian and Polish: A lot of words without or barely any vowels. To be honest I can't really tell which language they are. Just that they're from somewhere in the Eastern bloc. I tend to say Russian because of the 'hard' G.”

“Russian-harsh sounding, lack of vowels”

“Russian. Harsh consonants.”

“Russian- stress and articulation of consonants”

“Polish; The same as with Ukrainian: A lot of words without or barely any vowels. To be honest I can't really tell which language they are. Just that they're from somewhere in the Eastern bloc.”

“I think it's Czech or an Eastern European language Because of the way the speaker makes some vowel sounds in the back of her mouth”

“Ukrainian; a lot of words without vowels. Mostly consonants.”

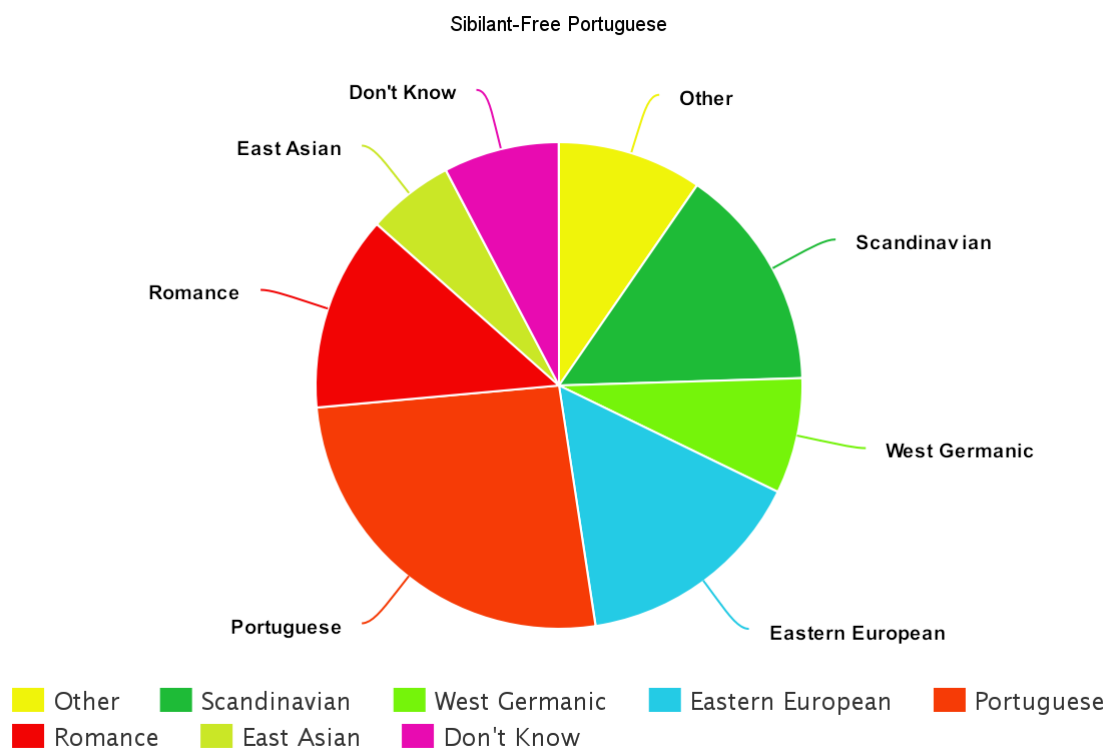
“I haven't got a clue, I can here loads of latin-based words, but there are a lot of harsh consonants too. Will say Serbo-Croat but I don't really believe it's that.”

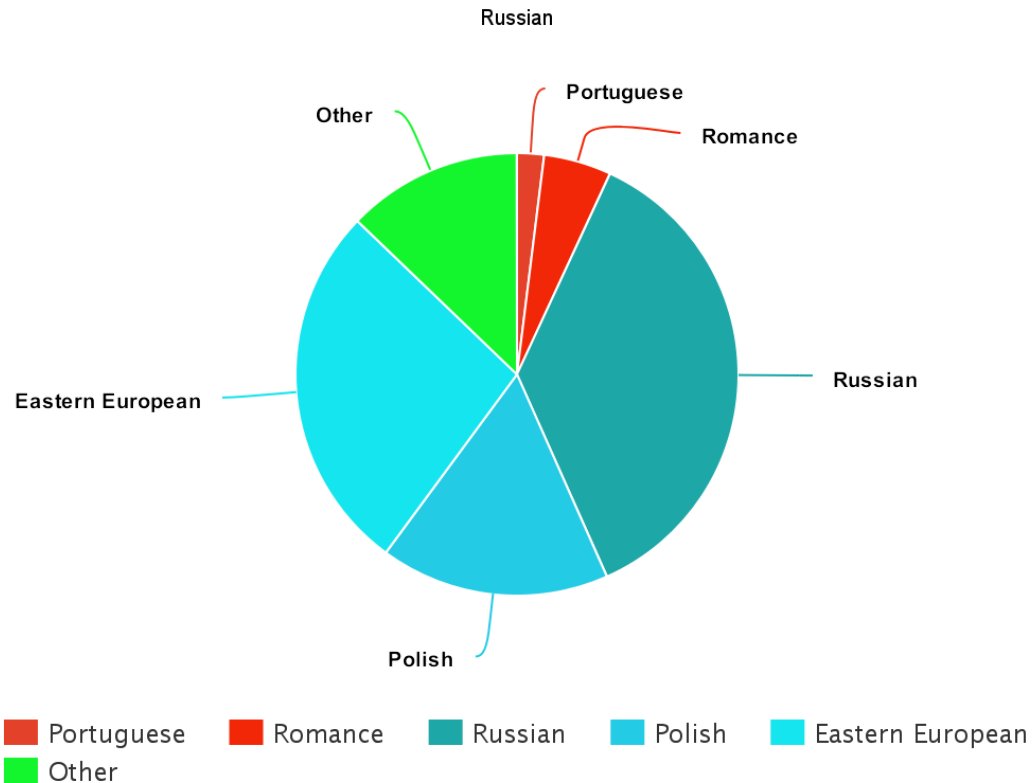
“I can't tell, I want to say an eastern European one due to the accent and the hard consonants.”

“Some easter european language, probably from the south (Bulgaria etc). I think so because of the strong consonants and the mildnes of vowels”

It is interesting to note that the unfamiliarity of Brazilian Portuguese as a Romance or European language does not cause people to pick Eastern European languages, presumably because it is not stress-timed enough, and it doesn't have enough consonants. (Frota and Vigário actually found it to be half-way between Spanish and French on the one hand and Japanese on the other, making it almost mora-timed!) This influenced people to be more likely to pick an Indo-Iranian language than an Eastern European one if they did not recognise it as a Romance language (22 people picked Arabic or Turkish and only 7 people picked Russian or Polish - the most commonly picked Eastern European languages). Most people believed it to be a Romance language; 152 to be precise. Despite being Portuguese's closest relative, only 19 people guessed Spanish, 26 guessed French, 10 Romanian, 7 Italian, and of course 86 Portuguese.

## 20. Results: Hypothesis Two: Does European Portuguese sound Eastern European because of its sibilants?





meta-chart.com

It was possible to use two controls for comparison for this hypothesis: European Portuguese without its sibilants removed (as opposed to excerpt three which did have them removed) and excerpt four which was Russian. Russian is thought to be a language with lots of sibilants. Having said that, only about 7 (out of 74) of the individual responses which guessed Russian, referred to its sibilants (and not all exclusively, the “ovs” and “nyets” for example point more to the uninvestigated palatal consonants hypothesis than hypothesis two):

“Russian Quick and a lot of ch's sounds”

“Russian or a country nearby, sounds kind of closed- mouth, lots os 'its' sounds”

“Russian, there were a lot of 'ch' sounds which I associate with this language.”

“Russian. Lot of "tszj" sounds”

“Russian? The "cz" sound makes me think of this.”

“Sounds like Russian, lots of "sk" sounds”

“Russian or polish sounded like a few 'ovs' & 'nyets' in there”

The table below clarifies the differences in perception between the two Portuguese excerpts and Russian:

	Portuguese	Eastern European
European Portuguese	38	137
EP no Sibilants	54	32
Russian	4	163

With only 32 people guessing an Eastern European language for Portuguese with no sibilants, clearly removing the consonants, if not removing the sibilants, made it seem less Eastern European. To be thorough, an odds ratio test was performed with both controls, as can be seen below:

	Eastern European	Not Eastern European
European Portuguese	137	63
EP no Sibilants	32	168
Russian	163	37

Odds ratio with 95% confidence interval

	Estimate	Lower	Upper
Portuguese	1.00000	NA	NA
No Sibilants	11.29641	7.051542	18.55535

	Estimate	Lower	Upper
Russian	1.00000	NA	NA
No Sibilants	22.77783	13.72321	38.96825

P value

	midp.exact	fisher.exact	chi.square
Portuguese	NA	NA	NA
No Sibilants	0	$7.796313e^{-42}$	$3.119584e^{-39}$

	midp.exact	fisher.exact	chi.square
Russian	NA	NA	NA
No Sibilants	0	$2.461589e^{-27}$	$2.197329e^{-26}$

The p-value of 0 (rounded) illustrates that there is a miniscule probability of these results occurring by chance. In addition, the high slope figures of 11.3 and 22.8 show strong discrimination between the presence and absence of sibilants.

Unlike hypothesis one, hypothesis two is a very specific one to make, so much so that it encompasses hypothesis one as a result. The results could be explained, in increasing order of generality, by:

- The removal of sibilants
- The removal of fricatives
- The removal of consonants
- This removal of consonants makes the altered Portuguese seem less stress-timed as the  $\Delta C$  is lowered and the %V is raised

The possibility of the least general explanation - the removal of sibilants - being true is supported by this long list of individual responses which point to sibilants as the cue for picking an Eastern European language throughout the survey (many more people than those who mentioned features exclusive to stress-timing, such as "harsh consonants" or "lots of consonants"):

"Russian. There's a 'sh' sound you hear in languages in eastern europe."

"Russian again, I think I heard 'och' sounds which I associate with Russian"

"Russian, there were a lot of 'ch' sounds which I associate with this language."

"Russian? The "cz" sound makes me think of this."

"Russian/slavic, there is a vowel sound that sounds russian and there is a 'zsh' sound that sounds russian"

"Russian - the shhh sound at the end of certain words."

"Russian. Because it has a lot of dark 'L"s and zwzws sounds."

"Russian. The use of shka and ž"

"Russian, dj and sh sounds"

"Russian. The gsh and csh sounds on the ends of words."

"Polish - "Sh" sounds and inflections which sound the way Polish speakers speak English."

"Polish, I think I've heard this before when I lived in Aberdeen. There is a 'sh' sound I heard a lot which made me think of that."

"Polish, the solid sounds of the words. Lack of melody in the speech pattern. Harsh s sounds"

"Polish. Predominance of z sounds and sounds using "front of mouth"

"Polish - the 'such' sound is similar to those I've overhead Polish people speaking."

"Polish. The sounds- sh"

"Czech, from the intonation, and lots of shhh sounds"

"Icelandic - it sounded quite Norse, the oer (u) sound. Lots of sht sounds. Could be Latvian?"

"Hungarian? Completely unintelligible to me but many 'sh' sounds?"

"Albanian. I've only met a couple of guys who ever spoke it but the way they pronounced the "isht" sound is very similar."

"Eastern European - it's got that back of the mouth sound with lots of 'sh' in it"

Some of the orthographically transcribed sounds could be velar fricatives rather than sibilants particularly as some of them were from excerpt five where the perceived velar fricative turned out to be a rather significant cue. It is also worth noting that quite a few participants used the sibilant as a cue for Portuguese and not Spanish when they heard Romance sounding words. Apparently the sibilants of Portuguese stand out in the language,



which drives some people to think of Eastern European languages, and allows some people to recognise it as Portuguese:

“Portuguese - some a bit like Spanish but with the 'sh' sound”

“Portuguese - similar word endings to Spanish, but with more added in mouth sounds (sh is very common I believe?)”

“Portuguese, I have studied it and understand it. The sh and zh sounds make me think it's from continental Europe”

“Portugese/spanish? "s" sounds”

“Portuguese - it sounds like a Latin language with a lot of fricatives Russian - the tone of the ladies voice, the shh sound”

“Portuguese - similar 'sh sh' noises (this one's from Polish)”

“Portugese I understood some words but this is not Spanish and the 'sh' sounded portugese”

“Portugese I understood some words but this is not Spanish and the 'sh' sounded portugese 'pollo' is spanish but also portugese”

The reason why the excerpt three pie chart is more complicated than the other ones, is that the lack of the sibilant cue threw a lot of participants off and created a whole host of other cues, the main one being tone.

According to David James Silva (1994):

“EP is characterized by the following three properties: (1) stress that is clearly marked by an increase in pitch and volume, (2) a significant number of vowel reduction processes, and (3) a tolerance for complex syllable onsets and codas.”

Perhaps cues two and three were removed along with the sibilants because apparently tone was much more apparent once sibilants were removed from the transcript. One could point to the cutting and pasting of speech segments leading to strange intonation if it were not for the fact that it was excerpt one, unaltered European Portuguese - unaltered in the sense that neither nasal vowels nor sibilants were removed - was the only recording subject to this kind of manipulation. Excerpt three was cut and paste before the recording and not after, so the content was nonsensical, but it still should have read normally.

Norwegian is a language that is well known for being tonal and Scandinavian languages were selected by 31 people for this excerpt. To put that into perspective, a Scandinavian language was selected no more than five times for any other excerpt and it was selected 46 times in total (so excerpt three made up two thirds of the total). The individual justifications specifically mentioning tone as a cue for picking a Scandinavian language are listed below:

“Swedish. The lilt”

“Finnish - a Nordic language I've never heard. The 'l' was very thin, the au and eu sounded Nordic. The musicality (higher and lower notes) of the words, less rhythmical. But it also sounded a bit southern european too.”

“I just don't know! It had the sing song lilt of a Scandinavian language or something like Dutch but none of the fricatives you'd expect, it also had some of the syllable structure of maybe a Latin language - I don't know”

“scandinavian or norse. sounds a bit more lilting than russian, but also some germanic crunchiness.”

There are only four examples of the above but a few more claiming Scandinavian languages adamantly without any other phonological justifications, and Norwegian and Swedish are set apart from the rest of Europe (except for Limburgian and some dialects of Latvian) as using pitch to convey differences in meaning:

“Norwegian Sounds stereotypically Scandinavian”

“Sounds like someone for Norway or Sweden”

“Sounds Nordic. Swedish? Another guess.”

“Swedish, Danish, Norwegian or Finnish. I thought it sounded Scandinavian.”

“Swedish or Russian (sounds like annoyingly loud tourists on a bus)”

“Finnish. It sounds Nordic since there are a lot of really strange vowel sounds.”

“Finnish, sounds Scandinavian but not Danish or I think Norwegian.”

Similarly 16 West Germanic languages were selected - a couple of them explicitly for tonal reasons - which has similarly rare selection except for excerpt five where the perception of a velar fricative turned out to be quite a big cue, and in this excerpt that velar fricative cue (which turns out to be a uvular trill, but that will be explained later):

“Dutch. Because it is very lyrical like Scandinavian languages however it doesn't quite match them. It is pronounced with a bit of a Flem sound to it. There are also a few other pronunciations that sound more southern European either a bit French-like "modiale" or also a bit Portuguese.”

“Dutch. Sounded like it might be, the speaking pattern seemed Dutch, rising and then softly falling on some words.”

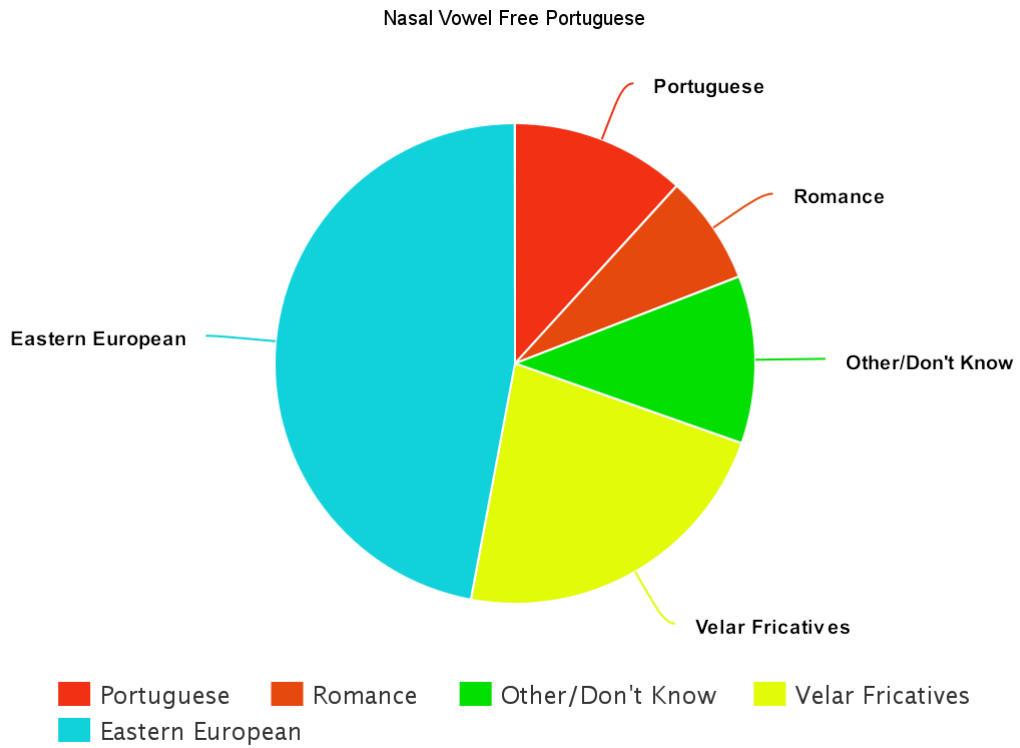
Finally, what makes excerpt three stand out is that so many people chose “other” languages. 16 languages to be exact, and the other five excerpts only generated 10 between them, with a maximum of three for any individual one. 13 of these languages belong to the oriental or East Asian group, and to put that into perspective, the other five excerpts had three between them. The justification for this a couple times was intonation (although this is ironic for Japanese which is very flat in comparison).

“Japanese because of the intonation.

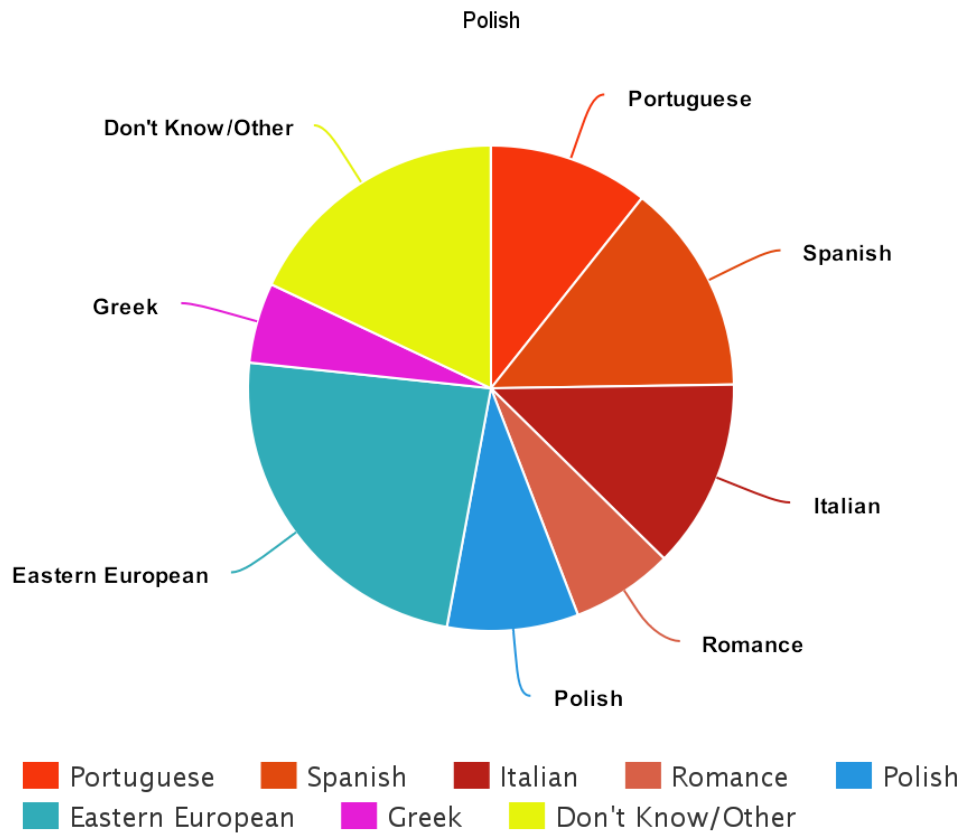
Japanese. Oriental, varying tones on the same sound and repetition of short words, but didn't sound like a Chinese language.”

Vietnamese was also included three times and never justified but Vietnamese is stereotypically tonal. The proportionally high nomination of these languages for excerpt three, languages which were not chosen when listening to the other excerpts, could be due to tone, as many people suggest, or perhaps the listeners were confused by the sibilants having been removed, in which case it would be the case that sibilants are a cue for perceiving an Eastern European language... but also a Romance one, so hypothesis two cannot be confirmed.

**21. Results: Hypothesis Three: Does European Portuguese sound Eastern European because of its nasal vowels?**



meta-chart.com



meta-chart.com

Of the three hypothetical methods; removing stress-timing, removing sibilants, and removing nasal vowels, clearly the latter had the least effect with 96 people still perceiving Portuguese to be Eastern European. As the table below demonstrates, there is still an effect, whether this is because of removing nasal vowels or not is yet to be confirmed:

	Portuguese	Eastern European
European Portuguese	38	137
EP no Nasals	24	96
Polish	22	67

Do Portuguese's nasal vowels make it sound Eastern European? According to this table, removing them made it slightly less Eastern European sounding, and it also made Portuguese slightly less Portuguese sounding. What is interesting is that Polish, which is an Eastern European language which uses more nasal vowels than most, was perceived as Eastern European fewer times than Portuguese without nasal vowels. Already this implies that nasal vowels aren't an Eastern European cue to foreigners. It also solidifies the main hypothesis that Portuguese sounds Eastern European; the number of participants who thought Portuguese was Eastern European was more than double the number who thought Polish was Eastern European. Before doing a statistical analysis of the effect of nasal vowels, it is worth noting that any effect was perceived implicitly, as below is the entire list of individual responses which pointed at nasal vowels as a cue for anything, and two out of three of those contradict hypothesis three:

"Polish. I am sure it is a Slavic language but it doesn't sound like the Balkan languages that I have heard. Doesn't sound harsh enough or nasal enough to be Russian. So Polish seems like a reasonable guess!"

"Portuguese, because of the nasal vowels (and it's Judite)"

"Russian because it sounds Eastern European but not as nasal as some others"

The results of the odds ratio are as follows:

	Eastern European	Not Eastern European
European Portuguese	137	63
EP no Nasals	96	104
Polish	67	133

Odds ratio with 95% confidence interval

	Estimate	Lower	Upper
Portuguese	1.00000	NA	NA

No Nasals	2.348728	1.565942	3.545551
-----------	----------	----------	----------

	Estimate	Lower	Upper
Polish	1.00000	NA	NA
No Nasals	0.5468886	0.3638871	0.8181752

## P value

	midp.exact	fisher.exact	chi.square
Portuguese	NA	NA	NA
No Nasals	3.300159e <sup>-5</sup>	4.694109e <sup>-5</sup>	3.22486e <sup>-5</sup>

	midp.exact	fisher.exact	chi.square
Polish	NA	NA	NA
No Nasals	0.003269527	0.004316715	0.00316811

There is very high significance, but not as high as for the first two hypotheses with p values that aren't exactly 0 and the p value of the nasal vowel to Polish isn't that much lower than 0.05. The comparisons although showing significance, are contradictory, with the Portuguese to Portuguese odds ratio showing that nasal vowels cause Portuguese to sound less Eastern European (if they show that at all) and nasal vowels causing Polish to sound less Eastern European than Portuguese without them (if they show that at all). The slopes of 0.55 and 2.35 are also a lot lower than the ones previously found in this experiment. Clearly the "significance" of the results are driven by many other factors.

Hypothesis three has been voided, but what is causing the uniqueness of the results of nasal vowel free Portuguese and Polish? One finding these final two excerpts has at least confirmed, is that Portuguese sounds so Eastern European that it sounds more Eastern European than an Eastern European language itself. A significance test to confirm this for good measure is on display below:

	Eastern European	Not Eastern European
European Portuguese	137	63
Polish	67	133

Odds ratio with 95% confidence interval

	Estimate	Lower	Upper
European Portuguese	1.00000	NA	NA
Polish	4.293906	2.836507	6.568981

P value

	midp.exact	fisher.exact	chi.square
European Portuguese	NA	NA	NA
Polish	1.919798e <sup>-12</sup>	3.1338e <sup>-12</sup>	2.534165e <sup>-12</sup>

The p value isn't exactly 0 and the slope of 4.29 isn't as big as the results deemed significant so far but these results still safely show that European Portuguese is perceived, statistically significantly, as more Eastern European than Polish.

The pie chart for nasal vowel free Portuguese has a yellow slice labelled "velar fricative"; this label is for all the languages which aren't Romance or Eastern European and which have velar fricatives; around 50. The "velar fricative cue" as it is now being coined, could be what is taking away (or potentially adding to) the Eastern European slice. Excerpt five is the only excerpt, certainly of Portuguese, that has any "guttural" sounds. The sound being referred to on this occasion is the uvular trill / / which is found at the beginning of words in European Portuguese and in geminates.

Orthographically speaking, it is found where there is an "r" at the beginning of a word or where there is a double "r". Surprisingly, the other two Portuguese excerpts feature the uvular trill only once between them, whereas excerpt five features it five times (in the words "correr", "repetida", "rítmico", "corredor", and "roda"). Perceptually, to speakers who use guttural sounds in their first language, the uvular trill and the velar fricative sound the same, perhaps because the trill is further back than the fricative and the fricative is stronger than the trill. (The Spanish alveolar trill is perceived as less similar to the uvular trill than the uvular trill is to the velar fricative, at least to Spanish, English, and Brazilian Portuguese speakers). As such, the following individual responses to excerpt five point to these sounds as being the cue for their selection of languages:

"Russian. Guttural sounds I associate with that language"

"Russian. Harsher and some echts at the end of some words"

"Ukrainian, because it sounds like Russian but more guttural."

"Serbia, sounds slavic. Very throat based sounds"

"Wow. It has dark L's but also the hard uvular G (like in Dutch). So, maybe, Georgian? Or Armenian?"

"Difficult! Sounds a bit like Arabic (the guttural H sounds) or Russian - the guttural L sounds and the soft zh sounds. Therefore I guess Romanian because it has been influenced by both the Slavic and Arabic languages."

“also sounds like an Eastern European language, but not Russian, the 'ch' sounds are unfamiliar to what I heard of Russian, but also 'thick l', not sure!”  
 “Hebrew, because of the sound of the g/ch”  
 “Ivriet G-sound”  
 “Hebrew. Based on the way the letter g is pronounced”  
 “Hebrew. It contains the hard 'ch' sound but it's not German Dutch or Gaelic”  
 “Hebrew. The accent is very phlegmy which makes me think it is Middle Eastern in origin so I'm guessing Hebrew.”  
 “Arabic or similar due to the guttural tones”  
 “Israeli. Lots of throaty 'sh' and 'ch' sounds.”  
 “Arabic? It sounds Eastern European but there is X a velar fricative used...”  
 “Actually this might be ARabic or Egyptian...it has that catch in the back of the throat on some words”  
 “I'm torn between German and Arabic. arabic because it's guttural and German because it sounds too close to Latin to be Arabic”  
 “German as think I heard "Ich"?”  
 “German - sounds guttural”  
 “German as I heard 'sch' sound”  
 “Dutch - strong 'ch' sound”  
 “Dutch, the g sounds”

Below are some individual responses to excerpt five which confirm the belief that even without nasal vowels, Portuguese sounds Eastern European (and also the mention of what could be uvular trills wrongly misinterpreted as “glottal stops” rather than velar fricatives since “glottal stop” could imply any kind of guttural constriction considering the place of articulation):

“Ukrainian. Like Russian but rakes of glottal stops”  
 “Russian - glottal sounds and inflection similar to Russian people speaking English”  
 “Bulgarian? Sounds slavic. Lots of glottal stops”  
 “Russian? Sounds pretty hard and definitely Slavic in origin so possible that it could be Russian”  
 “Also sounds Russian”  
 “Russian it sounds like james bond baddie (sorry that's a stereotype type)”  
 “Ukraine. Sounds russian, but not as Russian as the previous one”  
 “Ukranian, it seems to share similarities with Russian but I was unable to pick the exact language and Ukranian is the most similar to Russian as I can think,”  
 “slawic language as well... guessing ukrainian”  
 “Uzbekistani sounds a bit Russian ish”  
 “Also sounds vaguely russian, could be Romanian or Czech”  
 “Don't know. Eastern European ... Don't think Polish but could be.”  
 “Another Slavic language maybe Polish”  
 “Estonian - because it also sounds russian”  
 “Bulgarian Sounds angry”  
 “Georgia ? In between Arabic and Russian”  
 “Georgian. Has a slight russian Middle East style”

“Estonia. Or something eastern European”

“Latvian. A guess, somehow sounds right”

“Complete ignorance here but going to go Hungarian; sound like eastern European intonations but with a big mix of sounds otherwise.”

“Croatian- sounds similar to the last one and they are close by geographically”

It is interesting to note that although the three European Portuguese transcripts were picked at random, they still have phonological variety, excluding the artificial changes. Despite that fact, contrasting excerpts (such as one with no uvular trills and one with five) they are all still deemed to be Eastern European sounding, more than can be said for Polish. So why is Polish not deemed an Eastern European language? It could be that participants who heard Portuguese four times as well as Russian had by the end of the experiment, run out of Eastern European languages to guess. As with sibilant-free Portuguese, quite an assortment of languages were guessed. Although it is not a Romance language, a Romance language was guessed 91 times (22 Portuguese, 29 Spanish, and 26 Italian). A theory that perhaps this study supports is that nasal vowels don't make a language sound Eastern European, rather they make it sound Romance. The odds ratio test lent statistical support to null hypothesis three. Another reason could tie in with hypothesis one, in that Polish isn't as stress-timed as other Eastern European languages. Nespor (1990) showed that Polish and Catalan share roughly the same space on the rhythmic scale continuum; between a stress-timed and syllable timed language. Some of the individual responses justifying it as a Romance language or Greek (which was guessed 11 times out of 20 times on this excerpt) are below:

“Italian - I think I recognise a couple of words”

“Italian - Some kind of dialect??? Rhythms sound like a cross between French and Spanish”

“Italian I heard eh sounds”

“Italian as had some similar sounds but also some very different ones!”

“Italian, I'm not sure why though”

“Italian, fast and flowing with lots of vowels”

“Italian, i thought i heard noche.”

“Italian- the lilt which I think differentiates from Spanish language”

“Italian.”

“Italian. Bits of it sounded like Italian words, though other bits didn't so much.”

“Italian. I dont know. for the word "Cosa" said in the voice over.”

“Italian. I have friends who sound like this.”

“Italian? It's sounds more like singing than the Eastern European languages.”

“Latin - similarities to French/Spanish “

“Ancient Roman. Because it sounds kind of like a romance language but not one I recognise”

“A French dialect. I recognised the word "avec" Haha the computer doesn't.....”

“French - I heard words I understood”

“French - again the quickness of the words”

“French - sounds familiar”

“French. The was he pronounces some words sounds French”

“Sounds like french. I thought I recognised certain words.”



“Macedonian, soft meditaranian influences but still has eastern sounds. Some possibly more Middle eastern influences”

“Sounds slight Greek and Eastern European - Macedonian”

“Portugese. Sounded like Jose Mourinho, slightly Spanish, slightly Russian.”

“Portuguese - similar 'sh sh' noises”

“Portuguese .. The accent doesn't sound Russian!”

“Portuguese? Sounds Spanish like”

“Greek - sounded a bit Mediterranean but not Western Europe.”

“Greek again. Sounds like the first one I guessed was Greek”

“Greek. Because it sounds like a southern-European language but it isn't Spanish or Italian or Serbo-Croat.”

“Greek. Because it has a flow when spoken like Italian or Spanish but the pronunciations are distinctly different and i'm sure it is neither of them.”

“Greek. Don't know why”

“Greek? This is a total guess”

“Greek? The double s sound periodically.”

“Spanish Sounds Hispanic”

“Spanish Just the accent and pronunciation”

“Spanish - guess”

“Spanish - it sounded more familiar”

“Spanish, some of the words sound Spanish, sounds quite fast”

“Spanish. It was quite free flowing, I also thought it might have been Italian but I wasn't sure. It wasn't as role off the tongue as I've heard Italian spoken before.”

“Spanish/Portuguese or similar, very rhyming and regular syllable structure and lots of recognisable sounds”

A few points are worth making again here, for a lot of people “the flow” of Polish made it sound Romance - particularly Spanish or Italian. The sound of it “flowing” probably comes from it being syllable timed and having a higher %V. More “flowing” perhaps means fewer constrictions caused by consonants. Secondly, as with hypothesis two, some individual responses really used sibilants as a cue for picking a language, most of them picked Eastern European languages, but it could also go the other way, particularly with Portuguese being picked, and so Portuguese being picked could have been because Portuguese sounds Eastern European in a sense, and a reason for this is its sibilants. Finally, Greek being picked meant Eastern European wasn't picked as often, the justification there was often that it didn't sound Western European but people didn't want to go so far to call it Eastern European because it's “not as harsh as Russian”.

One final point worth making is that while to most people, Portuguese sounds Eastern European, not everyone fell into the booby trap of this experiment. Undeniably, Portuguese's closest relative in Spanish, and although it was only guessed 23 times out of 600 (for the European Portuguese dialects), many people who had not heard Portuguese before or did not recognise it picked it because it resembled Spanish in some way, as can be seen from a handful of individual responses:

“Catalan - Couple words sounded similar to Spanish (ending sounds like gracias, buenas noches)”

“Spanish from somewhere in the world, maybe latin America! similar words the person say I recognize”

“Portuguese - Couple words sounded similar to Spanish (a partir de)”

“Portuguese? Sounds Spanish like”

“Portuguese - sort of sounds like Spanish- it's a guess”

“Portugese - mixture of spanish and french”

“I think Portugese, as it sounds like nothing I have ever heard (which is how some people describe Portugese) but I still detect some Spanish influences.”

“Portuguese sounded a bit French but bit Spanish”

“Spanish/Portuguese or similar, very rhyming and regular syllable structure and lots of recognisable sounds”

“Portuguese as don't think Spanish but some sounds think similar”

“Portuguese. Some of the words are similar to Spanish, and I remember the premise. The phonology sounds Slavic.”

“Portugese. Sounds somewhere in the middle of Spanish and Russian”

“Portuguese. Because it's quite deep, also particularly so because she's a woman. There are a few words that sound Spanish or very similar to Spanish. And there was at least one word that ended with an "aish" sound which is a common pronunciation in Portuguese for words ending in an "s".”

“Portuguese. Seems like they are speaking Spanish a little but then not completely because Spanish is a little more easier to pronounce.”

“tricky! some things sound like Spanish, but are pronounced as if with an Eastern European slant (thick l, for example), so I guess Portuguese?”

“Portuguese. Sounds similar to Spanish..”

“Portuguese. Sounds like a rough mixture between Easter European languages and Spanish.”

“Portuguese. Pronunciation and intonation. Similar to Spanish.”

## **22. Discussion**

Although there were three hypotheses as to why European Portuguese is perceived as Eastern European, the main one was that it is stress-timed. This is what I believe sets it apart the most from the other Romance languages. To approach this from the typical participant of this survey, perhaps one who is monolingual and hasn't paid much attention to foreign languages before, the main distinction they would draw with, for example, Russian and Spanish is that Spanish sounds like it has a lot of vowels and Russian sounds like it barely has any. Although they won't know isogloss theory by name they will most likely be well aware of it, and so Europe going from west to east gets more “consonanty”, or in other words, stress-timed. And indeed, generally speaking, this is true with the Germanic languages sandwiched by the Balto-Slavic and Romance languages sitting medially on the isochrony scale as well as the geographical one. In a sense, the subjects of the experiments have done what small children do when they learn a language: recognise a pattern and overgeneralise. European Portuguese is an exception to the rule. If Portugal were geographically located in the same place as it finds itself on the stress-timing continuum, it

would find itself a lot closer to the Iron Curtain. Portugal is the most western country in Europe - but it doesn't sound like it.

The other two hypotheses are of course important to mention, a summary of all the results is given in the table below:

Excerpt	% guessed Eastern European	% guessed Romance
European Portuguese	69	25
Brazilian Portuguese	13	76
EP no sibilants	26	40
Russian	81	7
EP no nasals	48	20
Polish	33	45

The results of hypothesis two are quite convincing, if we were to factor out hypothesis one, or if hypothesis one were invalidated then certainly hypothesis two would be very difficult to refute. As it were, hypothesis two cannot be disconnected from hypothesis one, since H1 is such an all encompassing hypothesis. Of course, the upside of this study is that it contains quantitative and qualitative data, and there is a significant amount of qualitative data which explicitly mentions sibilants being a cue for perceiving an Eastern European language. Given the results, and with hindsight, if the same experiment were to be repeated then the hypothesis I would run would not only be specific enough to be separable from hypothesis one but this specificity may also carry with it more accuracy. Hypothesis Four: European Portuguese is perceived as Eastern European due to its palato-alveolar sibilants. Sibilants are present in both Spanish and Portuguese but only Portuguese contains palatalised ones and it is perhaps this that makes it sound different from Spanish, and also this that makes it sound Russian or Polish, which have many palatalised consonants. The fact that hypothesis three was refuted comes as no surprise. I was quite sure I would find some support for hypothesis one before I started the experiment, and not so with hypothesis three. H3 is also at odds with H1. While H1 stresses the importance of vowel reduction, H3 points to vowels as a significant cue for both Eastern European languages and Portuguese, H1 and H3 are almost by nature contradictory. In fact, H3 acted more as a control for H1 than it did as a genuine hypothesis.

### **23. Limitations to the Study**

Despite the conclusions which have been drawn in the preceding discussion, it is important to consider any limitations to the study which could have affected the results.

There were two main drawbacks with the methodology. The first of which was definitely more of a technical error in that the last two excerpts weren't recorded with the highest quality and sounded quite tinny due to the recorders not speaking close enough to the mic. This problem was quashed by upping the intensity of the two Wav files but only after half the participants had taken part.

The second drawback was less easy to solve and hence wasn't. Many subjects were led to believe that there would be six different languages since they weren't told otherwise - some even claimed to have felt tricked - and so perhaps didn't opt for the same answer more than once when with more explanation, they might have done. At a stretch it could be argued that this may have created a positive bias for the results. With four languages being the same, to come up with different languages, participants might have had to come up with different Eastern European languages since many people know what most of the Romance languages sound like. Still, it probably would have been unwise to explicitly say "some languages might be the same" as it would be giving too much information away. It would be interesting to repeat this study with the ambiguity removed to check whether there was indeed a positive bias.

Although a lot of statistical analysis was undertaken - with more than one control in most cases, it was not very in depth. The main method of displaying the results were through pie chart ratios displaying what participants guessed the languages were, rather than an in-depth statistical analysis of what these results meant. The statistical analysis only served to prove that the results were significant and did not do much more than that. In fact doing much more than that with the methods used would have been counter productive. All in all, the study was probably limited by its statistical analysis and a different method would probably have been more effective. Although combining a quantitative and qualitative study and its data can be thorough and informative, it does make it difficult to analyse concisely. One final limitation to the study is that next to none of the background literature or research was in Portuguese which is the language being investigated, so the current study's foundation is limited in scope and depth.

## 24. Conclusion

137 out of 200 people believed European Portuguese was an “Eastern European” language. Probably a very similar number of people believe it is a Balto-Slavic language to state this as a more accessible fact. That is over two thirds. Of course, this is not true for Spanish or Portuguese speakers, but the fact that such high results in favour of the main hypothesis were discovered despite some native speakers being involved, it is not exaggerating to claim that at least two thirds of people hearing Portuguese without much prior knowledge of it or of Spanish, will believe it to be a language of somewhere between the Berlin Wall and the eastern border of Europe.

Two reasons were found for this. One more salient and credible than the other, but nonetheless two reasonably good claims: European Portuguese is an anomaly as a stress-timed Romance language, it is therefore not recognisable as a Romance language, nor is it recognisable as a Germanic language as most English speakers have a pretty good idea what they sound like, nor does it sound so alien that it could be perceived as a non-Indo European language. That leaves one alternative: a Balto-Slavic or similarly sounding language, and these ones in particular because they are stress-timed. They have consonant interval variability and vowel to consonant ratios, which sound like European Portuguese. The second reason being that European Portuguese is full of sibilants, as are languages like Russian and Polish. On the way, a number of other discoveries were made but not confirmed, which perhaps could be confirmed empirically in the future:

- European Portuguese sounds more Eastern European than Polish
- Portuguese without sibilants sounds tonal
- Uvular trills sound like velar fricatives and this can make Portuguese sound “guttural”.

## References

Abercrombie, David. (1967): Elements of General Phonetics. Edinburgh U.P. Chicago: Aldine.

Alexander, C. Norman Jr (1967): Portuguese Vowels in the United States } Social Forces 45. p278-81.

Azevedo, Milton M. (2005): Portuguese: A linguistic introduction Cambridge University Press

BBC (2016): Reality Check: Which EU countries contribute most to UK population? <http://www.bbc.co.uk/news/uk-politics-eu-referendum-36522985>

Bain, Read (1928): Portuguese Vowels in the United States } American Journal of Sociology 33. 940-57.

Carvalho, Joana (2012): Portuguese Vowels in the United States } Europeu Faculdade de Letras da Universidade do Porto.

de Carvalho, Joaquim Brandão (2011): Portuguese Vowels in the United States } Revista de Estudos Linguísticos da Univerdade do Porto - Nº 1 - Vol. 6 p51-66

Cohen, Rip. (2010): Portuguese Vowels in the United States } 500 Cantigas d' Amigo: Edição Crítica / Critical Edition <https://jscholarship.library.jhu.edu/bitstream/handle/1774.2/33843/8-Sketch%20of%20the%20Historical%20Phonology%20of%20Galician%20Portuguese.pdf>

Crystal, David (2010): The Cambridge Encyclopaedia of Language Cambridge University Press. Third Edition.

Dauer, Rebecca. (1983): Portuguese Vowels in the United States } Journal of Phonetics 11:51-62.

Ellison, Fred P. and Andrews, Norwood. (1969): Portuguese Vowels in the United States } In D. D. Walsh (Ed.) A Handbook for Teachers of Spanish and Portuguese. Lexington, MA

Entwistle, William (1953): Aspects of Language. Faber & Faber London.

Feuerlicht, Ignace. (1955): Portuguese Vowels in the United States } American Speech, Duke University Press. Vol. 30, No. 3

Fishman, Joshua A. & Agheyisi, Rebecca (1970): Portuguese Vowels in the United States } Anthropological Linguistics, Vol. 12, No. 5 pp. 137-157

Frota, Sónia. And Vigário, Marina. (2001): *Á } Á@Á | | ^ | æ ^ Á - Á @c@ æ Á ã ç & ç } • Á@Á Ò : | ] ^ ç ð æ ã ç Á } | ç \* ^ ^ Á æ ^ Probus 13.2, pp.247-275*

Holt, Eric (2016): *Ø | { Á æ Á } Á } | ç \* ^ ^ Á æ Á } | | \* æ Á @ ç } \* ^ ^ Á In João Costa, Sergio Menuzzi & W. Leo Wetzels, (eds) The Handbook of Portuguese Linguistics, . Wiley-Blackell.*

Hopper, Paul. and Traugott, Elizabeth Closs (2003): Grammaticalization. 2nd edition. Cambridge University Press.

Ihemere, Kelechukwu Uchechukwu (2006): *Ø Á ç \* | æ ^ á Á } | | æ ç Á @ Á ç á ^ Á - Á Š ç \* ^ æ ^ Á ç á ^ Á ç á Á ç \* ^ Á Á ç á ^ Á @ Á ç ^ Á - Á @ Á, ^ | | Á - Á } | ç æ ç | ç Ó ç University of Newcastle upon Tyne, United Kingdom  
<http://www.lingref.com/cpp/acal/36/paper1424.pdf>*

Jassem, Wiktor (2003): *Ú | | ã @ Journal of the International Phonetic Association 33. 103–107.*

Jebesen, H. and Biel, Steven H.(1986): Let's Go: The Budget Guide to Spain, Portugal and Morocco. New York: St. Martin Inc. Harvard Student Agencies

Jensen, John B. (1989): *U } Á @ Á ~ ç æ Á ç | | ã ç Á - Á } ç ã ç Á } | ç \* ^ ^ Á Source: Hispania, Vol. 72, No. 4pp. 848-852. American Association of Teachers of Spanish and Portuguese*

Jesus, Luis M. T. and Shadle, Christine H. (2002) *Ø Á ç æ ç ^ ç á ç á ^ Á - Á } | | ^ ç æ ç & ç æ ç | ç æ ç Á - Á } | | ^ ç Á } | ç \* ^ ^ Á æ ç ^ Á • Journal of Phonetics 30, 437–464*

King, P. D. (1972): Law and Society in the Visigothic Kingdom Cambridge University Press. p1-5

Ladefoged, Peter; Johnson, Keith (2010), A Course in Phonetics (6th ed.), Boston, Massachusetts: Wadsworth Publishing

Nespor, Marina. (1990): *Á } Á @ Á @c@ Á æ ç ^ ç | Á } | | \* ^ Ë In Iggy Roca (ed.) Logical Issues in Language Acquisition, , 157-175. Dordrecht: Foris.*

Padgett, Jaye and Zygis, Marzena (2003): *V @ Á ç | | ç } Á - Á ç ç • Á } | | ç á Á ~ • ç In: Hall, T.A. & S. Hamann (eds.). ZAS Papers in Linguistics 32, 155-174.*

Pike, Kenneth L. (1945): The intonation of American English. Ann Arbor: University of Michigan Press.

P.T.W. and E.L. The Economist (2012): The “East” is dead. Time to scrap “Eastern Europe”  
<http://www.economist.com/blogs/easternapproaches/2012/07/%E2%80%9Ceast%E2%80%9Ddead>

Ramus, Franck. Nespor, Marina. and Mehler, Jacques. (1999): *Journal of Linguistics* 73: 265-292.

Silva, David James (1994): *Journal of Linguistics* 30: 1-20. in Susan C. Herring & John C. Paolillo, eds. *UTA Working Papers in Linguistics* 1

Trudgill, P. and Giles, H. (1983). *Journal of Linguistics* 19: 1-20. In P. Trudgill (ed.), *On Dialect*. Oxford: Blackwell.

Walter, E. and Pronzato, L. (1990): *Journal of Linguistics* 26: 195-213. *Journal of Linguistics* (Journal of IFAC) archive Volume 26 Issue 2, Mar. 1990 Pages 195-213