Game Theoretic Models of Clear versus Plain Speech

by

Jie Jian

B.Sc. (Statistics), Huazhong University of Science and Technology, 2016

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Mathematics
Faculty of Science

© Jie Jian 2018
SIMON FRASER UNIVERSITY
Fall 2018

Copyright in this work rests with the author. Please ensure that any reproduction or re-use is done in accordance with the relevant national copyright legislation.

Approval

Name:	**Jie Jian**
Degree:	**Master of Science (Applied and Computational Mathematics)**
Title:	**Game Theoretic Models of Clear versus Plain Speech**
Examining Committee:	**Chair:** John Stockie Professor
	Paul Tupper Senior Supervisor Professor
	Yue Wang Supervisor Professor
	JF Williams Internal Examiner Associate Professor
Date Defended:	**November 27, 2018**

Abstract

Clear speech is a speaking style intended to improve the comprehension of the hearer, which is usually due to the external noise, less ideal listening conditions, or the speaker is intended to be more intelligible. Clear speech, which exhibits increased duration, pitch, amplitude, and more exaggerated articulation, consumes more energy in order to improve the likelihood of accurate communication. To strike a balance between the cost of clear speech and the improvement it brings, we use game theory to model the phenomenon of clear speech. The conventions that speakers and hearers use to communicate are considered as equilibria in the communication game, and we need to make predictions of how the equilibria changes under the different circumstances. How our models correspond to what is experimentally observed, and what predictions are made for experimental results are discussed in the thesis.

In the basic model, we study the case where the speaker has to send one of two messages equally likely in one-dimensional acoustic space. Next, we make a further discussion of the basic model in a priori probability of the sent message, the number of messages, and the conflicts between clearness and comprehensibility. The third contribution of this thesis is to extend the one-dimensional acoustic space to two dimensions, by introducing uncontrastive and contrastive features.

Keywords: game theory; communication; phonetics

Acknowledgements

I would like to express my profound gratitude to my supervisor Professor Paul Tupper, for all his infinite patience and guidance. He is a real mathematician, pure scholar and perfect supervisor. I am extremely lucky to be his student. I could never ask for a better supervisor.

A very special thanks goes to Professor Yue Wang and Mr. Keith Leung for their unconditional support in providing important advice and assistance with their Linguistics background.

Last but not least, I want to thank my family and friends, for their love and support.

Table of Contents

Approval ii

Abstract iii

Acknowledgements iv

Table of Contents v

List of Tables vii

List of Figures viii

1 Introduction 1

1.1 Game Theory and Linguistics 1

1.2 A Brief Introduction of Phonetics 2

1.3 Motivation 3

1.4 Hypothesis 4

1.5 Chapter Summary 5

2 Basic Game Theoretic Models of Communication in Adverse Conditions 6

2.1 Signal Detection Theory 6

2.2 Modelling Communication Game 8

2.3 Optimizing the Payoff Function 12

2.4 Results for Basic Game Theoretic Models 13

2.4.1 Dependence of the Optimum on Effort Parameter k 13

2.4.2 Optimum Depending on Noise Amplitude σ 14

3 Further Discussion of Basic Game Theoretic Model 16

3.1 Different Probabilities for Different Messages 16

3.1.1 Optimum Depending on Effort Parameter k Given $\rho_a = 0.2$ 17

3.1.2 Optimum Depending on Noise Amplitude σ Given $\rho_a = 0.4$ 18

3.1.3 Optimum Depending on ρ_a 19

3.2 Four-Message Model 19

3.3 Conflicts Between Clearness and Comprehensibility 23

4 Multiple Phonetic Variables **29**

4.1 Two Contrastive Features 29

4.1.1 Decision Boundary is the Mid-perpendicular of Two Messages . . . 31

4.1.2 Analysis of the Optimum 33

4.1.3 Numerical Optimization 35

4.2 Uncontrastive and Contrastive Features 37

5 Conclusions **43**

Bibliography **45**

List of Tables

Table 1.1 44 English Phonemes, including 20 vowels and 24 consonants [16] 3

Table 2.1 Possible Outcomes on a Trial of a Yes-No Experiment. 8

List of Figures

Figure 2.1 Sketched figure of the signal and noise probabilities distributed within a decision-maker. 7

Figure 2.2 A schematic showing our basic model. The speaker is required to communicate one of two messages: a or b. They select signal values x_a or x_b which they transmit to the hearer. Noise in the communication channel leads to the hearer receiving a perturbed signal which they classify as either a or b based on the criterion c. 8

Figure 2.3 $F(x)$, the cumulative distribution function of a standard normal random variable. 10

Figure 2.4 $g(x)$, the function describing the cost in our model of emitting a signal with a given phonetic variable. 11

Figure 2.5 Optimum depending on effort parameter k. *Left:* The value of the optimal x_a, x_b and c for varying k and $\sigma = 0.05$. *Right:* The value of the maximized payoff $E(x_a, x_b, c)$ for varying k and $\sigma = 0.05$. . . . 14

Figure 2.6 Optimum depending on noise a amplitude σ. *Left:* The value of the optimal x_a, x_b and c for varying σ and $k = 0.05$. *Right:* The value of the maximized payoff $E(x_a, x_b, c)$ for varying σ and $k = 0.05$. . . . 15

Figure 3.1 The optimal gap curve. The x-axis represents the difference between the a priori probability of message b and message a, *i.e.* $\rho_b - \rho_a$. The red line represents the distance between x_a and c, and the blue line represents the distance between x_b and c. When $\rho_a = \rho_b$, $\Delta_a = \Delta_b$ which means the x_a and x_b are symmetric about c. When $\rho_a > \rho_b$, $\Delta_a > \Delta_b$. When $\rho_a < \rho_b$, $\Delta_a < \Delta_b$. 18

Figure 3.2 Optimum strategy and payoff depending on k given $\rho_a = 0.2$. *Top Left:* The value of the optimal x_a, x_b and c for varying k and $\sigma = 0.05$. *Top Right:* The value of the optimal x_a, x_b and c for varying σ and $k = 0.05$. *Bottom:* The value of the maximized payoff $E(x_a, x_b, c)$ for varying k and $\sigma = 0.05$. 19

Figure 3.3 Optimum strategy and payoff depending on σ given $\rho_a = 0.4$. *Left:* The value of the optimal x_a, x_b and c for varying σ and $k = 0.05$. *Right:* The value of the maximized payoff $E(x_a, x_b, c)$ for varying σ and k. 20

Figure 3.4 Optimum strategy and payoff depending on ρ_a given $\sigma = 0.05$ and $k = 0.05$. *Left:* The value of the optimal x_a, x_b and c for varying ρ_a. *Right:* The value of the maximized payoff $E(x_a, x_b, c)$ for varying ρ_a. 20

Figure 3.5 The four-signal model. *Top Left:* The value of the optimal x_a, x_b, x_c, $x_d(solidlines)$ and c_{ab}, c_{bc}, c_{cd} (dashed lines) for varying k and $\sigma = 0.05$. *Top Right:* The value of the maximized payoff $E(x_a, x_b, x_c, x_d, c_{ab}, c_{bc}, c_{cd})$ for varying k and $\sigma = 0.05$. *Bottom Left:* The value of the optimal x_a, x_b, x_c, x_d (solid lines) and c_{ab}, c_{bc}, c_{cd} (dashed lines) for varying σ and $k = 0.05$. *Bottom Right:* The value of the maximized payoff $E(x_a, x_b, x_c, x_d, c_{ab}, c_{bc}, c_{cd})$ for varying σ and $k = 0.05$. 22

Figure 3.6 Optimum in four-message model depending on σ. When a signal with value 0.4 has been received, it would be classified as message a if the hearer thought the signal was sent under $\sigma = 0.01$ and message b if $\sigma = 0.05$. 24

Figure 3.7 The optimal strategies for the speaker and the hearer in the four-signal model with two levels of noise. The condition is indicated along the x-axis. The y-axis indicates the value of the signal used for each of the four message by the speaker (in blue), and the three criterion points used by the hearer (in red). 27

Figure 4.1 A schematic showing the communication game in two dimensions. The speaker is required to communicate one of two messages: a and b. They select signal values (x_1, y_1) and (x_2, y_2) which they transmit to the hearer. Noise in the communication channel leads to the hearer receiving a perturbed signal in the 2D vowel space which they classify as either a or b based on the decision boundary c. 30

Figure 4.2 Mid-perpendicular of two points A and B 31

Figure 4.3 Probability density functions of two signals. Suppose the noise in the 2D model is a standard multivariate Gaussian random variable. The optimal decision boundary to decode the received signals is the perpendicular bisector to segment the speaker's strategies. 32

Figure 4.4 Payoff function in 2D constractive model. *Top Left:* The value of the payoff function given $m \neq n$. *Top Right:* The value of vector field of the payoff function given $m \neq n$. *Bottom Left:* The value of the payoff function given $m = n$. *Bottom Right:* The value of vector field of the payoff function given $m = n$. 36
Figure 4.5 Optimum strategies for speaker in 2D contrastive model, given $m = n$. For a given communication environment (with noise level σ and effort parameter m and n fixed), the optimum strategies for the speaker would be two points on a circle which are symmetric about the origin. The radius of the circle is determined by the communication environment. 37
Figure 4.6 Numerical optimum of 2D constractive features model given $m = n$. *Top Left:* The value of the optimal radius for varying σ given $m = n = 3$. *Top Right:* The value of the maximized payoff for varying σ given $m = n = 3$. *Bottom Left:* The value of optimum radius for varying m given $\sigma = 0.05$. *Bottom Right:* The value of the maximized payoff for varying m given $\sigma = 0.05$. 38
Figure 4.7 The optimal strategies for the speaker in 2D contrastive model given $m \geq n$. The solid line represents the optimal positions of the speaker's strategy varying m. When $m = n$, the two messages are located on any positions on the solid circle that are symmetric about the origin. Increasing m, optimal (x_1, y_1) and (x_2, y_2) move in the opposite directions on the x-axis to separate from each other. 39
Figure 4.8 Numerical optimum of 2D constractive features model given $n < m$. *Top Left:* The value of the optimal x for varying σ given $m = 3$ and $n = 2$. *Top Right:* The value of the maximized payoff for varying σ given $m = 3$ and $n = 2$. *Bottom Left:* The value of optimum x for varying m from 3 to $+\infty$ given $\sigma = 0.05$. *Bottom Right:* The value of the maximized payoff for varying m from 3 to $+\infty$ given $\sigma = 0.05$. 40
Figure 4.9 $\tilde{P} = 1 - e^{\frac{x}{\sigma}}$, the function describing the probability of uncontrastive feature. 41
Figure 4.10 Optimum strategy in 2D model depending on σ. *Left:* Optimum for 2D model with uncontrastive and contrastive feature. *Right:* Payoff depending on σ. 42

Chapter 1

Introduction

Game theory is the study of cooperation and conflict, which provides a structure to model the situations where strategies of several agents are involved. The focus of this thesis is to use game theory as a framework to analyze the difference between plain and clear speech quantitatively. This chapter will give a brief overview over the concepts of game theory and linguistics, and basic phonetics. We will introduce our motivation to establish the model in the thesis, and provide the hypotheses in the model.

1.1 Game Theory and Linguistics

Game theory was first introduced as a field in its own by mathematicians von Neumann and Morgenstern in their publication, *Theory of Games and Economic Behavior*, in 1944 [26]. In the 1951, John Nash demonstrated that there was always an equilibrium point in any game with a finite set of actions in his famous paper *Non-Cooperative Games* [27]. Since then, game theory has been widely applied in war, politics, psychology, economics, sociology, and biology, for it provides a mathematical model for the study of decision-making in a game where players have conflicting or mutual interest and make their strategies based on other players'. A social interactive decision situation is known as a *game*, and the decision maker involved in the game is referred to as a *player*. Each player in the game is driven by a well-formulated goal, which is formalized as the player's *payoff function*. Each player has a set of actions to select. After each individual action has been selected, payoff function assigns to the corresponding outcome a *payoff* or *utility* to each player. The players aim to optimize their own payoff by selecting actions from the possible strategy set. How to determine among the actions in the strategy set is subject of game theory [11].

An interactive decision-making game distinguishes between two situations. In the first situation, players are not able to make any binding agreements and each decision-maker acts independently from all other decision-makers, which is known as *non-cooperative game*. The second type of game is known as *cooperative game*, which allows players to write binding contracts and multiple decision makers can act as a group [9, 25]. Evolutionary game theory

describes game models in which players choose their strategies through a trail-and-error process. The players learn over time that some strategies work better than others [33].

As a framework with internal consistency and mathematical foundations, game theory becomes an efficient tool for modeling communication. In 2008, Gerhard Jäger demonstrated a framework considering communication as a game in his paper *Applications of Game Theory in Linguistics* [12]. Suppose there are two players in the communication game. One player, sender S, has private information about some type or event t in some finite set of events T. The other player, receiver R, does not know the knowledge about the event. Successfully passing the information from S to R is in the interest of both players. A strategy for S is to map from events set T to a set F containing finite number of signals, from which S can transmit one signal. After observing a signal, R is able to make a guess about what event S has sent. R can map F to T as a tactic. Coalitions are formed between S and R, since a correct guess will give both players positive utility, otherwise they obtain nothing. Therefore, both S and R need to create best strategies respectively in order to maximize the possibility for R to hit a correct guess.

1.2 A Brief Introduction of Phonetics

Phonetics is a branch of linguistics dealing with the physical reality of speech sounds, where the human sounds in general are studied without reference to their systemic role in a specific language [5]. Phonetics can be divided into many areas, with the three main branches as *acoustic phonetics* which focuses on the transmission of the speech, *articulatory phonetics* studying the movements and actions of the speech organs in producing sounds, and *auditory phonetics* concerning how listeners perceive the sounds.

Articulation is the work of speech organs to generate speech sounds. *Manner of articulation* refers to the closure or constriction used when the sound is made. *Place of articulation* means the area in the mouth where the closure and constriction occur. Based on these two dimensions of articulation, namely place and manner, speech sounds can be classified as *vowels* and *consonants*. From view point of phonetics [28, 4], a *vowel* is a central-oral frictionless, voiced sound, produced with an open vocal tract. In different contexts, speakers are able to adapt their speech styles. Vowels can be further segmented into different sounds like $/i:/$, $/e/$, and so on, by various of phonetics features, mainly the first two formants F_1 and F_2 [10]. *The first formant frequency*, known as F_1, measures the vowel hight, which is also the frequency of the lowest characteristic resonance of the vowel. While *the second formant*, F_2 is the feature of backness to measure the degree of lip rounding. *Vowel space* is a two-dimensional area with respect to the first formant frequency F_1 and the second formant frequency F_2. *Pitch* refers to the fundamental frequency of phonation, since the perceived pitch has the similar range as the physical frequency of normal speech [10].

Phonemes are the different sounds in a language [5]. For example, there are 20 vowels and 24 consonants in English, which make up 44 sound phonemes shown in the Table 1.1 [5]. Some vowels are produced with greater muscle tension in the articulations than other vowels [6]. We call the former vowels (e.g., /i/, /e/, /u/, /o/) as *tense vowels* and the latter (e.g., /ɪ/, /ʊ/, /æ/) as *lax vowels*.

44 English Phonemes			
Vowels		Consonants	
Phoneme	Example	Phoneme	Example
/iː/	seat /siːt/	/p/	pull /pʊl/
/ɪ/	sit /sɪt/	/b/	bull /bʊl/
/e/	set /set/	/f/	ferry /'feri/
/æ/	cat /kæt/	/v/	very /'veri/
/ɑː/	match /mɑːʧ/	/θ/	think /θɪŋk/
/ɒ/	pot /pɒt/	/ð/	then /ðen/
/ʊ/	good /gʊd/	/t/	take /teɪk/
/ɔː/	port /pɔːt/	/d/	day /deɪ/
/uː/	food /fuːd/	/z/	zoo /zuː/
/ʌ/	much /mʌtʃ/	/s/	sing /sɪŋ/
/ɜː/	turn /tɜːn/	/ʃ/	show /ʃəʊ/
/ə/	collect /kə'lekt/	/ʒ/	pleasure /'pleʒə/
/eɪ/	take /teɪk/	/ʧ/	cheap /ʧiːp/
/aɪ/	mine /maɪn/	/dʒ/	jail /dʒeɪl/
/ɔɪ/	oil /ɔɪl/	/k/	case /keɪs/
/əʊ/	no /nəʊ/	/g/	go /gəʊ/
/aʊ/	house /haʊs/	/m/	my /maɪ/
/ɪə/	hear /hɪə/	/n/	no /nəʊ/
/eə/ or /ɛə/	air /ɛə/	/ng/	sing /sɪŋ/
/ʊə/	tour /tʊə/	/l/	love /lʌv/
		/r/	round /raʊnd/
		/w/	well /wel/
		/j/	young /jʌŋ/
		/h/	house /haʊs/

Table 1.1: 44 English Phonemes, including 20 vowels and 24 consonants [16]

1.3 Motivation

Clear speech is a style whose properties are modified with the intention of being more comprehensible, *i.e.* a more enunciated speaking manner involving a greater degree of speech articulator movement which leads to corresponding changes in acoustic features [20]. Speakers use clear speech in many contexts, including a noisy environment, when speaking to the hearing impaired or language learners, or merely when the speaker wants to enhance intelligibility [29]. Two features of clear speech are found through experiments. In [17], the experiments on 12 speakers who were producing clear speech in different contexts observe that the clear speech is not identical in these different situations, but it shares some similar features that differ from plain speech: longer duration, higher pitch, and increased amplitude. Another important feature of clear speech was seen in [22] which indicates that in

order to make distinguishing different phonemes easier for the listener, phonetic differences between phonemes are exaggerated.

We observe numerous different effects in a variety of studies as the evidence of the two features: the common features of clear speech in different contexts, and the exaggeration of phonetic differences between phonemes in clear speech. In the presence of noise, the speakers will alter their acoustic efforts during the communications in order to speak clearly, such as increasing the vocal intensity, glottal spectral slope, formant structures, or fundamental frequency. This natural phenomenon is known as Lombard effort [19]. For example, in [8] the experiments and statistical analysis show that the Lombard effect is found in spectral target, dynamic formant movement, especially second formant $F2$ as the speech intelligibility improvements in English vowels. The experiments in [31] also concluded that speakers increased duration and average speech amplitude with the noise increase. There is a need to establish a framework to study these effects in order to predict the differences between clear and plain speech.

Many observed and predicted modifications in clear speech are in tension of each other, which offers another reason to design a predictive theory and quantitative models of the differences between clear and plain speech. This observation is documented in [20]. In the study, clear and plain productions are compared with respect to three pairs of English tense-lax vowels. The experiments reveal that though the difference between the tense and lax vowel is more significant in clear speech than in the plain speech, the more extreme articulatory gesture in clear speech makes the vowel sound like some other one in plain speech, which will confuse the receiver. Further study of how speakers resolve these contradictory demands and what paradigms can be used to explore the issue will be needed.

1.4 Hypothesis

One main assumption in our model is that the speaker uses clear speech in order to enhance the probability of transmitting the information correctly. In some cases, acoustic changes do not help make the speech more intelligible. For example, speaking loudly and slowly does not contribute to the understanding of a foreign language. However, the Lombard effect still works in many cases [8, 19, 31]. In our model, we assume that the clear speech style is developed to improve the quality of communication, and we apply this conjecture to predict the clear speech styles.

Following [7] and [12] introduced in Section 1.1, we model the clear and plain speech styles with the Game Theory by assuming a speaker and a hearer are engaged in a communication game that they can play over and over again. The speaker need to communicate one of several distinct possible message to the listener. To transmit the message, the speaker is able to to select a continuous-valued signal to emit. To make it easy for the receiver to distinguish the various messages, the speaker will assign different values of signal to different

messages. However, there is noise in the communication channel distorting the signal being conveyed in the channel. Therefore, the receiver only gets a perturbed signal and needs to apply a strategy to decode the received signal.

There are two important factors in our game. One is that the cost varies among different signals. The other is that the possibility of transmission of the wrong message issues from the presence of noise in the channel. As introduced in [22], H & H theory indicates that the speaker needs to find a balance between the cost spent in the communication and the probability to receive correct messages.

In our model, the space of messages T is discrete, as it is in the work of G. Jäger in [13]. According to the work in [1] and [14], we assume the signal of one message is continuous-valued.

1.5 Chapter Summary

This thesis is organized as follows: in Chapter 2, we introduce our basic game theoretic models to describe the communication game, the analytical solution to the basic model under different conditions, and numerical results from MATLAB. In Chapter 3, we further study the basic models based on three different conditions respectively: first, unequally likely messages; second, four-messages rather than two; and third, different noise patterns. In Chapter 4, we expand the basic model from a single phonetic variable to multiple phonetic variables, and introduce the concepts of contrastive and uncontrastive features. Parts of the Chapter 2, like Section 2.2, 2.3, and 2.4 are included in the paper [32].

Chapter 2

Basic Game Theoretic Models of Communication in Adverse Conditions

2.1 Signal Detection Theory

Signal Detection Theory (SDT) is one of the most successful mathematical models in cognitive science. It was developed in the early 1950s, and it provides the strategy to evaluate the probability of making a mistake when you make one decision based on the phenomenon observed [23]. For example, a doctor needs to interpret CT images and detect whether or not there is a tumor. The tumor can be considered as a signal, and the task for the doctor is to decide whether the signal is present. If the doctor diagnoses the sample as a tumor, then there is a chance of making a mistake, while there will be a possibility of miss if the sample is diagnosed as no tumor. In this case, the doctor has to decide between reporting a tumor or not. Signal Detection Theory is widely applied in such uncertain or ambiguous situations where an individual must make decision whether or not some condition is present. Although the basic decision is a simple alternative, such as *diagnosing as tumor* or *not tumor* in the above example, the incomplete and random information makes the decision difficult. There are always chances to make erroneous choice, no matter how intelligent the decision maker is. We will introduce Signal Detection Theory, which was developed for this type of task, to detect a weak signal occurring in a noisy environment, through the basic detection experiment known as *"Yes-No" Design* [30].

In the basic detection experiment, the observer will face two kinds of signals. First, a trial without any systematic component and only the random background environment is presented, referred to as the Noise trials. Some signal is added to the Noise trial, which forms the other trials, called signal plus noise trials, or, Signal trials. The observer will receive evidence either from the Signal or Noise with uncertainty. The value received is a random variable, denoted as X_S for the signal trials and X_N for the noise trial. Figure 2.1

shows the density functions $f_N(x)$ and $f_S(x)$ of the two random variables. We define the accumulative distribution functions of X_N and X_S as $F_N(x)$ and $F_S(x)$ respectively.

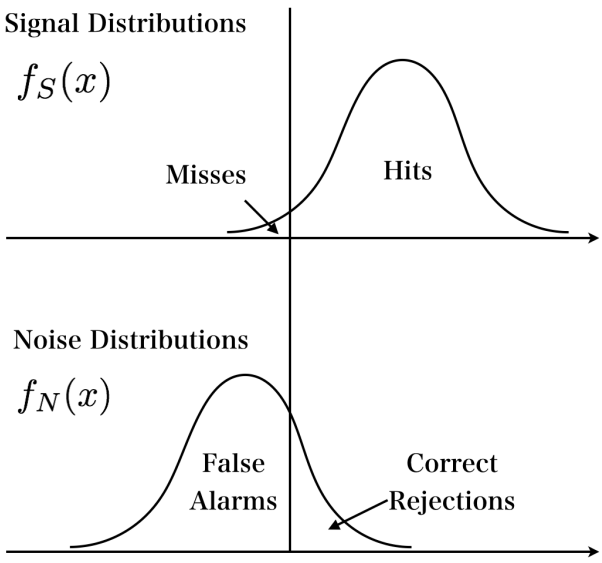

Figure 2.1: Sketched figure of the signal and noise probabilities distributed within a decision-maker.

Suppose the observer who is trying to distinguish the Signal and the Noise, is fully aware of the probability distribution of the two and has to set a response criterion c for decision making, sketched in Figure 2.1. The observer reports *"YES"* when the amount of evidence for the signal is larger than c and *"NO"* when it is smaller than c. Response *YES* to a Signal is known as a *hit*, and the error where the observer reported *YES* when there is only Noise is a *false alarm*. A *miss* is the case where *NO* has been reported when a Signal present, and a correct rejection is to reply *NO* to a Noise. Each type of response can occur with each type of trial, so the four possible outcomes are identified by name in Table 2.1.

The probability of a hit can be calculated as the area under the density function $f_S(x)$ above the threshold c, which can be written as

$$\begin{aligned} P_{Hit} &= P(\text{YES}|Signal) \\ &= P(X > c|Signal) \\ &= P(X_S > c) \\ &= \int_c^{+\infty} f_S(x)dx \\ &= 1 - F_S(c). \end{aligned} \tag{2.1}$$

Similarly, the probability of false alarm can be represented as

$$P_{False} = P(\text{YES}|Noise) = P(X_N > c) = \int_c^{+\infty} f_N(x)dx = 1 - F_N(c). \tag{2.2}$$

	Response	
Stimulus	YES	NO
Signal	hit	miss
Noise	false alarm	correct rejection

Table 2.1: Possible Outcomes on a Trial of a Yes-No Experiment.

2.2 Modelling Communication Game

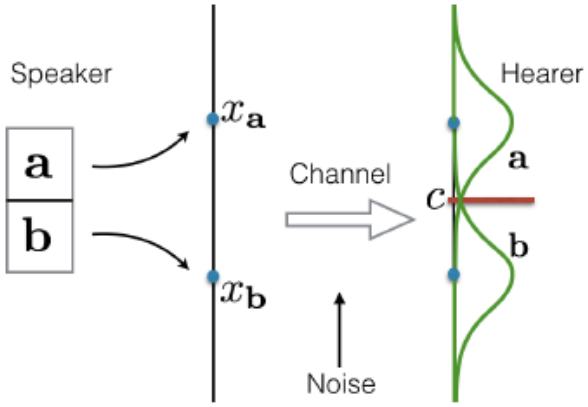

Figure 2.2: A schematic showing our basic model. The speaker is required to communicate one of two messages: a or b. They select signal values x_a or x_b which they transmit to the hearer. Noise in the communication channel leads to the hearer receiving a perturbed signal which they classify as either a or b based on the criterion c.

We present in this section the most basic version of our communication game model between two agents, the speaker and the hearer. There are two messages for the speaker to transfer to the hearer: either a or b. Each time the speaker can only convey one of the two messages, so the speaker acts one of two types, *sending message a* or *sending message b*. Which message the speaker has sent is not known to the hearer. The hearer must decide which message the speaker has released based on the received signal, which means the hearer has a choice between two actions: *diagnose the received signal as message a* and *diagnose the received signal as message b*. We assume that the consequences of mistakenly transmitting a for b are the same as for transmitting b for a.

In a hypothetical situation, we can imagine the speaker ordering a drink at a coffee shop, with a meaning "coffee", and b meaning "tea", and each choice being made equally often.

The choice will be expressed by the speaker via a continuous-valued variable, the value of which is set by the speaker. Thus, if the speaker is to emit a tone for a given length of time, they will be able to select the value assigned to each option - tea and coffee. In doing so, noise will be added, during the process like the production, transmission or reception of the tone, affecting the signal before it even reaches the hearer. The hearer therefore is tasked to decipher an already-noisy signal in order to extract the speaker's choice.

After the hearer gives an interpretation of the received signal, the corrective feedback will be given to both of the players. We assume that the game is played over and over again so that an optimal communication system is able to be developed. Both the speaker and hearer want to maximize the utility, and their goal is the same. The game is cooperative, then there is no confusion about which type of equilibrium should be reached since the two players can be engaged in the game as a team [7]. The strategy for the speaker is to select two values of the variable x, namely x_a and x_b, for the signals a and b respectively. Without loss of generality, assume that $x_a < x_b$. Suppose the noise in the communication channel is σn, where n a standard Gaussian random variable such that $n \sim \mathcal{N}(0,1)$, and σ is a noise amplitude. Therefore, the hearer will perceive either $y = x_a + \sigma n$ or $y = x_b + \sigma n$ depending on the message, a or b, the speaker sent. The received value y for the two messages obeys

$$x_a + \sigma n \sim \mathcal{N}(x_a, \sigma) \text{ and } x_b + \sigma n \sim \mathcal{N}(x_b, \sigma).$$

Based on the heard signal y, the hearer has to make a decision which message has the speaker sent. The hearer's task is an example of the standard model in Signal Detection Theory, introduced in Section 2.1. The receiver's optimal choice is to fix a value c as the criterion, and choose message a when $y \leq c$ and choose message b when $y < c$. In probability measure, the probability that $y = c$ is zero, so the case $y = c$ can be classified as either message a or message b. Following Section 2.1, the probability that the receiver hits the correct message a can be represented as

$$\begin{aligned} \mathbb{P}(\text{correct}|a) &= P(y \leq c) \\ &= P\left(x_a + \sigma n \leq c\right) \\ &= P\left(n \leq \frac{c - x_a}{\sigma}\right) \\ &= F\left(\frac{c - x_a}{\sigma}\right), \end{aligned} \tag{2.3}$$

where F is the cumulative distribution function of a standard normal random variable with mean 0 and variance 1, as shown in Figure 2.3 and can be written as

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt. \tag{2.4}$$

Similarly, the probability that the receiver hits the correct message b can be represented as

$$\begin{aligned}\mathbb{P}(\text{correct}|b) &= P(x_b + \sigma n > c) \\ &= 1 - P(x_b + \sigma n \leq c) \\ &= 1 - F\left(\frac{c - x_b}{\sigma}\right).\end{aligned} \tag{2.5}$$

Since there is no cost for the receiver to select the decision boundary, the optimal value of c will be the one that maximizes the probability of receiving the correct message. Since each message is equally likely, we can express this probability as

$$\begin{aligned}P(x_a, x_b, c) &= \mathbb{P}(\text{correct transmission}) \\ &= \mathbb{P}(\text{correct}|a)\mathbb{P}(a) + \mathbb{P}(\text{correct}|b)\mathbb{P}(b) \\ &= \frac{1}{2}F\left(\frac{c - x_a}{\sigma}\right) + \frac{1}{2}\left[1 - F\left(\frac{c - x_b}{\sigma}\right)\right].\end{aligned} \tag{2.6}$$

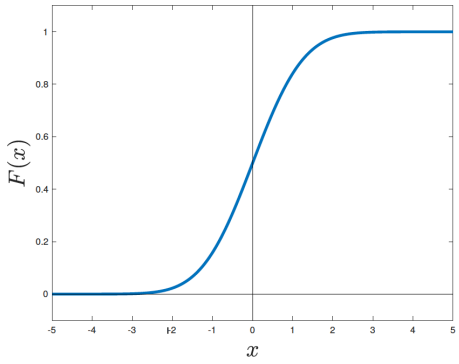

Figure 2.3: $F(x)$, the cumulative distribution function of a standard normal random variable.

If there are no penalty or cost for extreme x, the speaker will increase the gap between the messages as much as possible, since the probability of success will increase to 1 as the distance between x_a and x_b increases. In any realistic system there is either a finite range of possibilities for x, or there is a disincentive for using large or small values of x. The idea is that more extreme values of x require more effort, and the speaker will make less effort unless there is sufficient benefit to making more effort [22].

The probability model of correctly receiving the signals and the idea of higher cost in more exaggerated signals can be combined by defining a cost for emitting a signal that depends on x, the phonetic variable of the signal. Suppose that the effort required to emit signal x is $kg(x)$ where g is defined as

$$g(x) = \frac{1}{x(1-x)}, \text{ for } 0 < x < 1 \tag{2.7}$$

and set $g(x) = \infty$ for $x < 0$ or $x > 1$, as shown in Figure 2.4. k, which is called the *effort parameter*, is some positive constant we use to parameterize the overall effort in emitting a signal and it can evaluate the difficulty to expend effort. We chose this form for g for three reasons:

- The cost function (2.7) constrains only the sounds in the range $(0, 1)$ to be emitted and they all have positive cost;
- The cost for the more extreme sounds are higher which represents that they are more difficult to emit;
- Effort is close to constant for signals within the middle of range.

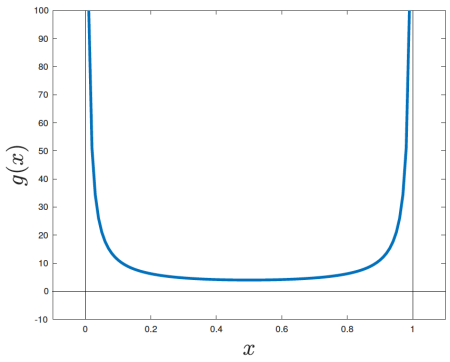

Figure 2.4: $g(x)$, the function describing the cost in our model of emitting a signal with a given phonetic variable.

Since the game is cooperative, we now make a fairly strong assumption for the purposes of simplicity: the speaker and the hearer have the same payoff function in the game. So they are equally interested in the correct message being transmitted, and are equally interested in the speaker's effort being minimized. This is clearly not always a reasonable assumption, and Chapter 4 will consider different models. Following this symmetric modelling choice, we assume that the expected payoff to the speaker and the hearer in one round of the communication game is

$$E(x_a, x_b, c) = P(x_a, x_b, c) - \frac{k}{2}\left(g(x_a) + g(x_b)\right), \tag{2.8}$$

that is, the probability of the message being correct minus the average cost to the speaker of transmitting x.

2.3 Optimizing the Payoff Function

We assume that the speaker and the hearer will choose the strategy that maximizes the value of the payoff function shown as (2.8), which can be written as:

$$\begin{aligned} E(x_a, x_b, c) &= P(x_a, x_b, c) - \frac{k}{2}(g(x_a) + g(x_b)) \\ &= \frac{1}{2}F\left(\frac{c - x_a}{\sigma}\right) + \frac{1}{2}\left[1 - F\left(\frac{c - x_b}{\sigma}\right)\right] - \frac{k}{2}\left(\frac{1}{x_a(1 - x_a)} + \frac{1}{x_b(1 - x_b)}\right) \\ &= \frac{1}{2}\int_{-\infty}^{\frac{c-x_a}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt + \frac{1}{2}\left[1 - \int_{-\infty}^{\frac{c-x_b}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt\right] \\ &\quad - \frac{k}{2}\left(\frac{1}{x_a(1 - x_a)} + \frac{1}{x_b(1 - x_b)}\right). \end{aligned} \tag{2.9}$$

To maximize the payoff function (2.9) of variables x_a, x_b and c, the technique of partial differentiation is used.

First, suppose x_a and x_b are fixed. The hearer tries to optimize his strategy based on the given x_a and x_b. Taking the partial derivatives of (2.9) with respect to c yields

$$\frac{\partial E}{\partial c} = \frac{1}{2\sigma}\left[f\left(\frac{c - x_a}{\sigma}\right) - f\left(\frac{c - x_b}{\sigma}\right)\right] \tag{2.10}$$

$$= \frac{1}{2\sigma}\frac{1}{\sqrt{2\pi}}\left[e^{-\frac{(c-x_a)^2}{2\sigma^2}} - e^{-\frac{(c-x_b)^2}{2\sigma^2}}\right]. \tag{2.11}$$

where f is the probability density function of the standard Gaussian distribution.

By the symmetry of Gaussian distribution and the assumption $x_a < x_b$, we know that $c = \frac{x_a + x_b}{2}$ is the only critical point of $\frac{\partial E}{\partial c}$. When $c < \frac{x_a + x_b}{2}$, a simple use of calculus shows that Function 2.11 $\frac{\partial E}{\partial c}$ is positive, and negative when $c > \frac{x_a + x_b}{2}$. The facts show that whatever x_a and x_b are, the optimal value for the criterion is $c = \frac{x_a + x_b}{2}$. So whatever the speaker chooses for x_a and x_b, the hearer will always choose the mid point as the criterion to gain the optimal payoff.

Define the equal gap $\Delta = c - x_a = x_b - c$. The basic models can be simplified as a double-variable problem by plugging $x_a = c - \Delta$ and $x_b = c + \Delta$ into the original model. Moreover, $F(\frac{\Delta}{\sigma}) = 1 - F(-\frac{\Delta}{\sigma})$ due to the symmetry of probability density function of normal distribution. The payoff function (2.9) can be rewritten as

$$
\begin{aligned}
E(x_a, x_b, c) &= \frac{1}{2}F\left(\frac{c-x_a}{\sigma}\right) + \frac{1}{2}\left[1 - F\left(\frac{c-x_b}{\sigma}\right)\right] - \frac{k}{2}\left(\frac{1}{x_a(1-x_a)} + \frac{1}{x_b(1-x_b)}\right) \\
&= \frac{1}{2}F\left(\frac{\Delta}{\sigma}\right) + \frac{1}{2}\left[1 - F\left(\frac{-\Delta}{\sigma}\right)\right] - \frac{k}{2}\left(\frac{1}{(c-\Delta)(1-c+\Delta)} + \frac{1}{(c+\Delta)(1-c-\Delta)}\right) \\
&= F\left(\frac{\Delta}{\sigma}\right) - \frac{k}{2}\left(\frac{1}{(c-\Delta)(1-c+\Delta)} + \frac{1}{(c+\Delta)(1-c-\Delta)}\right).
\end{aligned}
\tag{2.12}
$$

For fixed gap Δ, to maximize the above Payoff function (2.12) with respect to c, only the cost term $\frac{1}{(c-\Delta)(1-c+\Delta)} + \frac{1}{(c+\Delta)(1-c-\Delta)}$ needs to be minimized. Notice that $(c-\Delta)(1-c+\Delta)$ and $(c+\Delta)(1-c-\Delta)$ are positive since $0 < x_a, x_b < 1$. According to the fact that arithmetic mean is always larger than the harmonic mean, we have:

$$
\frac{1}{(c-\Delta)(1-c+\Delta)} + \frac{1}{(c+\Delta)(1-c-\Delta)} \geq \frac{2^2}{(c-\Delta)(1-c+\Delta) + (c+\Delta)(1-c-\Delta)},
$$

the equality holds if and only if $(c-\Delta)(1-c+\Delta) = (c+\Delta)(1-c-\Delta)$, *i.e.* $c = \frac{1}{2}$.

The symmetry of $g(x)$ about $x = \frac{1}{2}$ also implies that the optimum will always have $x_b - \frac{1}{2} = \frac{1}{2} - x_a$ implying $c = \frac{1}{2}$.

Hence, we only need to maximize the function $E(\frac{1}{2} - \Delta, \frac{1}{2} + \Delta, \frac{1}{2})$ with respect to $0 \leq \Delta < \frac{1}{2}$ to find the optimum of the original problem. If we assume the values for the noise amplitude σ and the effort parameter k, the specific optimal values for Δ can be solved numerically.

2.4 Results for Basic Game Theoretic Models

This optimization problems is solved by MATLAB's *fminsearch* routine. *fminsearch* was chosen since our optimization problem is unconstrained and nonlinear. In this section, we study how x_a, x_b and c depend on the noise amplitude σ and the effort parameter k using our computed solutions to the optimization problem.

2.4.1 Dependence of the Optimum on Effort Parameter k

Fixing the value of noise amplitude σ as 0.05, we increase the effort parameter k from 0 to 1. Figure 2.5 *left* shows how x_a, x_b and c depend on k for a fixed value of $\sigma = 0.05$. We see that as k goes to 0, x_a and x_b go to 1. The extreme values for x_a and x_b make sense, since in this limit, there is no penalty for making the gestures as large as possible and larger gap between x_a and x_b gives greater probability to recognize the correct signal. Likewise, as k goes to infinity, x_a and x_b both go to $\frac{1}{2}$, the cheapest possible signal, since the cost of emitting a signal becomes large compared to the benefit of accurate communication.

Figure 2.5 *right* shows that as k increases from 0 to 1, the optimized payoff keeps decreasing, since in the system it is getting harder to make effort.

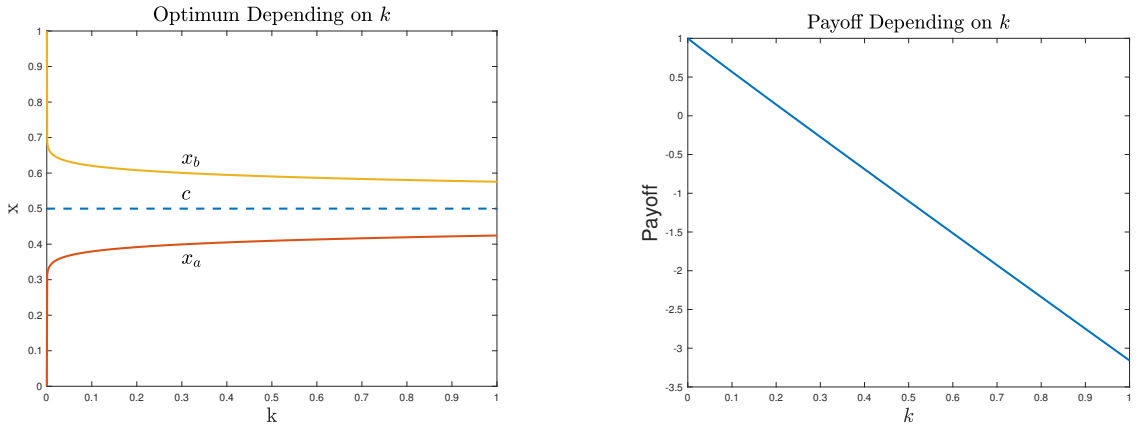

Figure 2.5: Optimum depending on effort parameter k. *Left:* The value of the optimal x_a, x_b and c for varying k and $\sigma = 0.05$. *Right:* The value of the maximized payoff $E(x_a, x_b, c)$ for varying k and $\sigma = 0.05$.

2.4.2 Optimum Depending on Noise Amplitude σ

The case of fixed k and varying σ is more interesting. Figure 2.6 *left* shows how x_a, x_b and c depend on σ for a fixed value of $k = 0.05$. For small values of σ, x_a, x_b go to $\frac{1}{2}$ as σ goes to zero as we might imagine. As when there is no noise, even the slightest difference between x_a and x_b gives perfect communication, and setting both to $\frac{1}{2}$ minimizes effort.

What happens as σ increases is less expected. Initially, as σ increases from 0, gestures become more extreme in order to improve the probability of correct communication. This is the standard clear speech effect, and is a key part of Lombard speech. The phenomenon agrees with the effects in $F2$ in English vowels [8] and duration and amplitude in [31].

What is striking is the case that after the signals past a certain noise level the effect reverses itself, and then become less extreme in our model. This occurs because, if the noise is large enough, the probability of communication regardless of the signals used is so low that it is no longer worth the effort to make the more extreme gestures that were worthwhile for a lower level of noise. We know of no observations of this phenomena, but predict that it will be observed for human subjects with sufficiently large amplitudes of noise. Indeed, [29] observes speech amplitude increasing with a decreasing rate as noise level is increased, and a reduction in amplitude may be observable if an even larger noise level is tried. A similar phenomenon has been observed in domestic fowl [2]. The chickens studied varied the frequency with which they repeated their calls in the presence of different amounts of noise. It was observed that for lower levels of noise the birds increased call frequencies with increasing noise, and it is conjectured that this is an adaption to improve the probability of communication by expending more effort. However, interestingly the authors noted that

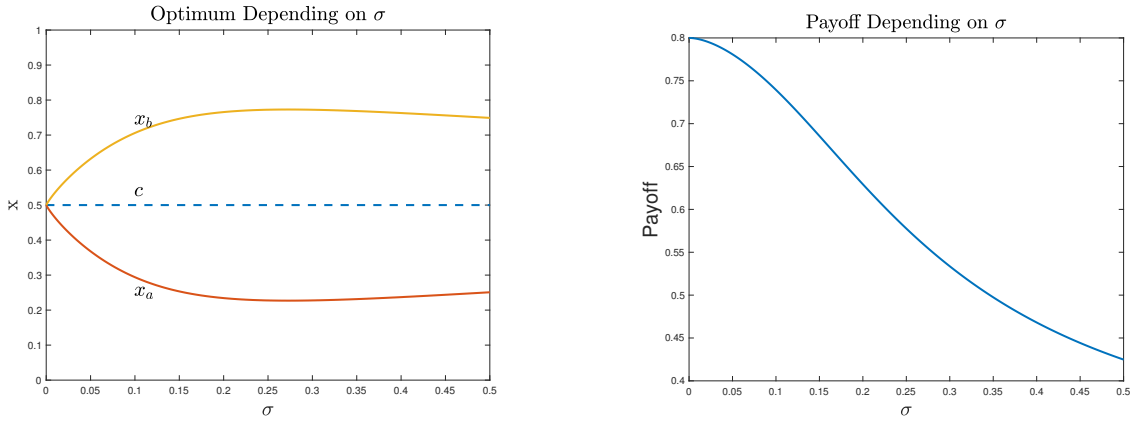

Figure 2.6: Optimum depending on noise a amplitude σ. *Left:* The value of the optimal x_a, x_b and c for varying σ and $k = 0.05$. *Right:* The value of the maximized payoff $E(x_a, x_b, c)$ for varying σ and $k = 0.05$.

after the noise was increased past a certain point, the birds decreased the frequency of their calls, as would be predicted by our model.

Figure 2.6 *right* shows that as σ increases from 0 to 0.5, the optimized payoff keeps decreasing, since the noise is increasing in the system.

Chapter 3

Further Discussion of Basic Game Theoretic Model

In this chapter, we will further discuss the basic game theoretic model in three aspects. First, we study how making a message more frequent changes its signal compared to the equally likely message model. Second, two more messages will be added to the basic model. Third, we will study how the signals will change if some of players are not aware of the noise level.

3.1 Different Probabilities for Different Messages

In Section 2.2, we make the assumption in the basic model that the two messages are transmitted equally often. That is to say we have the equal a priori probability for message a and b in (2.6): $\mathbb{P}(a) = \frac{1}{2}$ and $\mathbb{P}(b) = \frac{1}{2}$. However, this is not at all necessary for our model. In this section we will study how making a message more frequent changes the position of its signal in phonetic space.

We assume that the message a is transmitted with the prior probability ρ_a and b with ρ_b, where $\rho_a + \rho_b = 1$. Without loss of generality we assume that $\rho_a < \rho_b$. Hence, the probability of receiving the correct message is

$$\begin{aligned} P(x_a, x_b, c) &= \mathbb{P}(\text{correct}|a)\mathbb{P}(a) + \mathbb{P}(\text{correct}|b)\mathbb{P}(b) \\ &= \rho_a \cdot F\left(\frac{c - x_a}{\sigma}\right) + \rho_b \cdot \left[1 - F\left(\frac{c - x_b}{\sigma}\right)\right], \end{aligned} \tag{3.1}$$

and the cost of the communication is no longer the unweighted average of the costs of two messages. Instead it is the average weighted by the probability of the two messages,

expressed as $k\left(\frac{\rho_a}{x_a(1-x_a)}+\frac{\rho_b}{x_b(1-x_b)}\right)$. Hence, the payoff function is

$$\begin{aligned} E(x_a,x_b,c) &= P(x_a,x_b,c)-\frac{k}{2}(g(x_a)+g(x_b)) \\ &= \rho_a\cdot F\left(\frac{c-x_a}{\sigma}\right)+\rho_b\cdot\left[1-F\left(\frac{c-x_b}{\sigma}\right)\right]-k\left(\frac{\rho_a}{x_a(1-x_a)}+\frac{\rho_b}{x_b(1-x_b)}\right). \end{aligned} \tag{3.2}$$

Define the distance between x_a and c as Δ_a and the distance between x_b and c as Δ_b. Using the same technique as in Section 2.3, take the partial derivatives of Equation 3.2 with respect to c yields

$$\begin{aligned} \frac{\partial E}{\partial c} &= \frac{\rho_a}{\sigma}\cdot f(\frac{c-x_a}{\sigma})-\frac{\rho_b}{\sigma}\cdot f(\frac{c-x_b}{\sigma}). \\ &= \frac{\rho_a}{\sigma}e^{-\frac{\Delta_a^2}{2\sigma^2}}-\frac{\rho_b}{\sigma}e^{-\frac{\Delta_b^2}{2\sigma^2}}. \end{aligned} \tag{3.3}$$

If we set (3.3) to zero then we obtain

$$\Delta_b^2-\Delta_a^2=2\sigma^2\cdot\ln\frac{\rho_b}{\rho_a}>0. \tag{3.4}$$

Since we assume that $0<x_a\le x_b<1$, to make $\Delta_b^2-\Delta_a^2>0$ we must have $c<\frac{x_a+x_b}{2}$. For any given x_a, x_b and noise amplitude σ, the optimal c is the one satisfying equation (3.4). Fixing $\sigma=0.05$ and $k=0.05$, the curve of optimal Δ_a and Δ_b depending on $\rho_a-\rho_b$ is shown as the Figure 3.1. When the messages are equally likely sent, the optimal Δ_a and Δ_b are the same. We can see that if $\rho_a<\rho_b$ then $\Delta_a<\Delta_b$, which makes sense since it is worth it to share more space to the more frequent-sent message. In the following sections, we study how x_a, x_b and c depend on the noise amplitude σ, the the effort parameter k and the prior probability ρ_a for message a under the assumption that the messages are not equally transmitted.

3.1.1 Optimum Depending on Effort Parameter k Given $\rho_a=0.2$

We assume that the probability of the message a is 0.2 and then the probability of the message b is 0.8. Fix the value of noise amplitude σ as 0.05, we increase the effort parameter k from 0 to 1. Figure 3.2 *top left* shows how x_a, x_b and c depend on k for a fixed value of $\sigma=0.05$ and $\rho_a=0.2$. Similar to the purple lines when the messages are equally likely, as k goes to 0, x_a and x_b goes to 0 and 1 respectively; as k goes to infinity, x_a and x_b both approach to c.

Moreover, we compare the case of $\rho_a=0.2$ with the case of equally likely messages to find that the fact $\rho_a<\rho_b$ breaks the symmetry of the messages about $\frac{1}{2}$ and also rules out the invariance of the optimal strategy for the receiver. Just as the analytical solution equation (3.4) shows, $\Delta_a<\Delta_b$ when $\rho_a<\rho_b$. Yellow lines in Figure 3.2 *top left* demonstrates

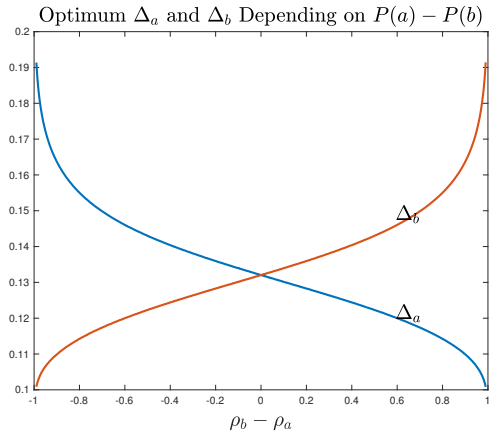

Figure 3.1: The optimal gap curve. The x-axis represents the difference between the a priori probability of message b and message a, *i.e.* $\rho_b - \rho_a$. The red line represents the distance between x_a and c, and the blue line represents the distance between x_b and c. When $\rho_a = \rho_b$, $\Delta_a = \Delta_b$ which means the x_a and x_b are symmetric about c. When $\rho_a > \rho_b$, $\Delta_a > \Delta_b$. When $\rho_a < \rho_b$, $\Delta_a < \Delta_b$.

that the bias of the prior probability pushes the optimal x_a and x_b toward the side of less frequent message x_a, which frees up more phonetic space and generates cheaper unit cost for the more-frequently used message.

For the multiple pairs of (Δ_a, Δ_b) given by changing k using our computed solutions, we plotted these numerical results to find that they agree with the analytical relation between Δ_a and Δ_b in Equation (3.4), shown as the *top right* subfigure in Figure 3.2.

Figure 3.2 *bottom* shows that as k increases from 0 to 1, the optimized payoff keeps decreasing. The payoff for the case $\rho_a = 0.2$ is slightly greater than the case where $\rho_a = 0.5$.

3.1.2 Optimum Depending on Noise Amplitude σ Given $\rho_a = 0.4$

Let the probability of the message a be 0.4, and then the probability of the message b is 0.6. Fix the value of effort parameter k as 0.05, we increase the noise amplitude σ from 0 to 0.5. Yellow lines in Figure 3.3 *top left* show how x_a, x_b and c depend on σ for a fixed value of $\sigma = 0.05$ and $\rho_a = 0.2$.

Similar to the purple lines when the messages are equally likely, as σ goes to 0, both x_a and x_b goes to $\frac{1}{2}$; as σ increases, x_a and x_b first separate away and then draw together after a certain noise level. Compared to the game when messages are equally sent, when we set ρ_a to 0.4 the receiver's strategy biases one way to the message a which gives more space to discriminant the potential message as message b, which agrees with (3.4) that as the noise level σ increase, the optimal c makes $\Delta_b^2 - \Delta_a^2$ increase.

Figure 3.3 *right* shows that as σ increases from 0 to 0.5, the optimized payoff keeps decreasing.

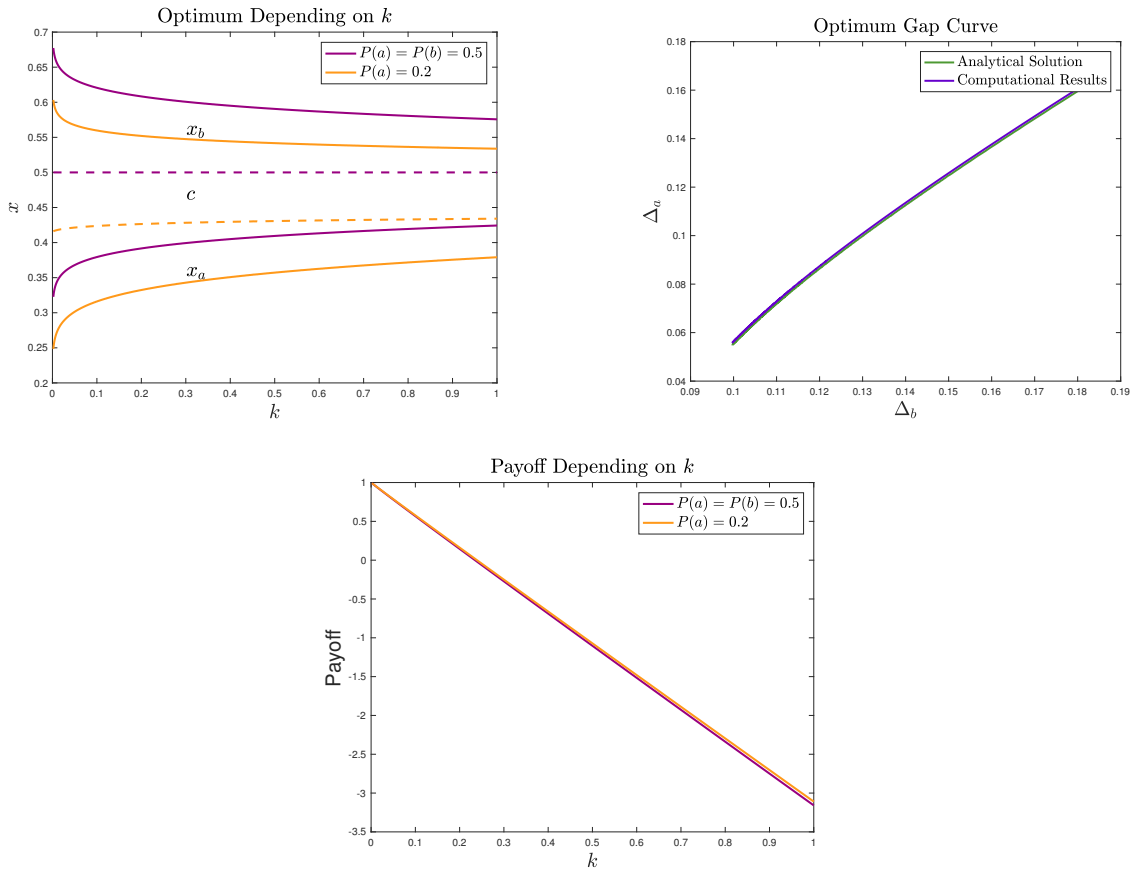

Figure 3.2: Optimum strategy and payoff depending on k given $\rho_a = 0.2$. *Top Left:* The value of the optimal x_a, x_b and c for varying k and $\sigma = 0.05$. *Top Right:* The value of the optimal x_a, x_b and c for varying σ and $k = 0.05$. *Bottom:* The value of the maximized payoff $E(x_a, x_b, c)$ for varying k and $\sigma = 0.05$.

3.1.3 Optimum Depending on ρ_a

Fixing noise amplitude σ to 0.05 and effort parameter k to 0.05, we increase the probability for message a from 0.005 to 0.5. Figure 3.4 *left* shows how x_a, x_b and c depend on ρ_a for a fixed value of $\sigma = 0.05$ and $k = 0.05$. As ρ_a increases, x_a, x_b and c increase. When $\rho = 0.5$, c becomes $\frac{1}{2}$ and x_a and x_b are symmetric about c.

Figure 3.4 *right* shows that as ρ_a increases from 0.005 to 0.5, the optimized payoff keeps decreasing, which indicates that the payoff in the single-message game is greater than the payoff in the two-message game.

3.2 Four-Message Model

The basic model in Chapter 2 can be extended to an arbitrary number of messages analogously. In this section we just consider the case of four messages which are equally likely. Suppose there are messages a, b, c and d in the message space, again corresponding to four

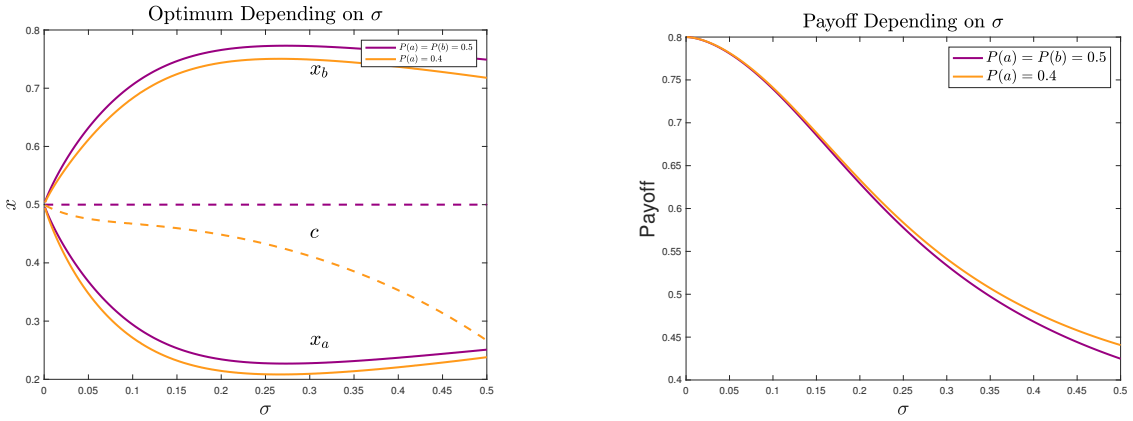

Figure 3.3: Optimum strategy and payoff depending on σ given $\rho_a = 0.4$. *Left:* The value of the optimal x_a, x_b and c for varying σ and $k = 0.05$. *Right:* The value of the maximized payoff $E(x_a, x_b, c)$ for varying σ and k.

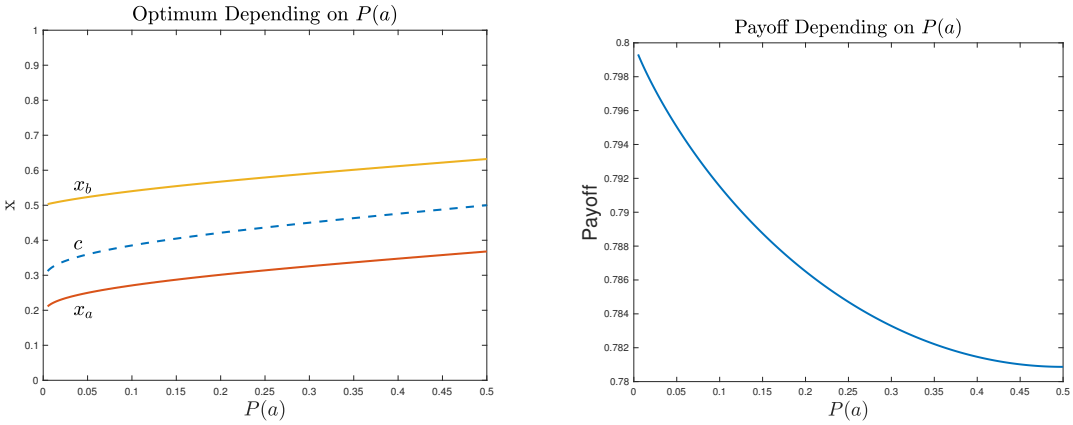

Figure 3.4: Optimum strategy and payoff depending on ρ_a given $\sigma = 0.05$ and $k = 0.05$. *Left:* The value of the optimal x_a, x_b and c for varying ρ_a. *Right:* The value of the maximized payoff $E(x_a, x_b, c)$ for varying ρ_a.

distinct meanings. The strategy of the speaker is to choose signals $x_a < x_b < x_c < x_d$ to represent the four messages, and the strategy of the hearer is to fix values c_{ab}, c_{bc}, and c_{cd}. When the hearer receives signal $y = x + \sigma n$, they select message a if $y \leq c_{ab}$, b if $c_{ab} < y < c_{bc}$, c if $c_{bc} < y < c_{cd}$, and d if $y > c_{cd}$. The expressions for the probability of

correct transmission are demonstrated in the following:

$$
\begin{aligned}
&P(x_a, x_b, x_c, x_d, c_{ab}, c_{bc}, c_{cd}) \\
&= \mathbb{P}\left(\text{correct transmission}\right) \\
&= \mathbb{P}\left(\text{correct}|a\right)\mathbb{P}(a) + \mathbb{P}\left(\text{correct}|b\right)\mathbb{P}(b) + \mathbb{P}\left(\text{correct}|c\right)\mathbb{P}(c) + \mathbb{P}\left(\text{correct}|d\right)\mathbb{P}(d) \\
&= \frac{1}{4}\mathbb{P}\left(x_a + \sigma n \leq c_{ab}\right) + \frac{1}{4}\mathbb{P}\left(c_{bc} > x_b + \sigma n > c_{ab}\right) + \frac{1}{4}\mathbb{P}\left(c_{cd} > x_c + \sigma n > c_{bc}\right) + \frac{1}{4}\mathbb{P}\left(x_d + \sigma n > c_{ab}\right) \\
&= \frac{1}{4}\left[1 - F\left(\frac{c_{ab} - x_a}{\sigma}\right)\right] + \frac{1}{4}\left[F\left(\frac{c_{ab} - x_b}{\sigma}\right) - F\left(\frac{c_{bc} - x_b}{\sigma}\right)\right] \\
&+ \frac{1}{4}\left[F\left(\frac{c_{bc} - x_c}{\sigma}\right) - F\left(\frac{c_{cd} - x_c}{\sigma}\right)\right] + \frac{1}{4}F\left(\frac{c_{cd} - x_d}{\sigma}\right),
\end{aligned}
\tag{3.5}
$$

Again the payoff function $E(x_a, x_b, x_c, x_d, c_{ab}, c_{bc}, c_{cd})$ is the the difference between the probability of correct transmission and the expected cost, shown as Function (3.6):

$$
E(x_a, x_b, x_c, x_d, c_{ab}, c_{bc}, c_{cd}) = P(x_a, x_b, x_c, x_d, c_{ab}, c_{bc}, c_{cd}) - \frac{k}{4}\left(g(x_a) + g(x_b) + g(x_c) + g(x_d)\right). \tag{3.6}
$$

Both speaker and hearer act to maximize the payoff function (3.6). Take the derivative of $E(x_a, x_b, x_c, x_d, c_{ab}, c_{bc}, c_{cd})$ with respect to c_{ab}, we get:

$$
\frac{\partial E}{\partial c_{ab}} = \frac{1}{4\sigma}\frac{1}{\sqrt{2\pi}}\left[-e^{-\frac{(c_{ab}-x_a)^2}{2\sigma^2}} + e^{-\frac{(c_{ab}-x_b)^2}{2\sigma^2}}\right]. \tag{3.7}
$$

By the symmetry of Gaussian distribution and the assumption $x_a < x_b$, we know that $c_{ab} = \frac{x_a + x_b}{2}$ is the only critical point of $\frac{\partial E}{\partial c}$. When $c_{ab} < \frac{x_a + x_b}{2}$, a simple use of calculus shows that $\frac{\partial E}{\partial c}$ is positive, and negative when $c_{ab} > \frac{x_a + x_b}{2}$. The facts show that whatever x_a and x_b are, the optimal value for the criterion is $c_{ab} = \frac{x_a + x_b}{2}$. So whatever the values of x_a, x_b, x_c, x_d, c_{bc} and c_{cd}, the optimal choice for the hearer is to use criterion points $c_{ab} = (x_a + x_b)/2$.

Similarly, taking the derivative of $E(x_a, x_b, x_c, x_d, c_{ab}, c_{bc}, c_{cd})$ w.r.t. c_{bc} and c_{cd} respectively, we get:

$$
\frac{\partial E}{\partial c_{bc}} = \frac{1}{4\sigma}\frac{1}{\sqrt{2\pi}}\left[-e^{-\frac{(c_{ab}-x_b)^2}{2\sigma^2}} + e^{-\frac{(c_{ab}-x_c)^2}{2\sigma^2}}\right], \tag{3.8}
$$

$$
\frac{\partial E}{\partial c_{cd}} = \frac{1}{4\sigma}\frac{1}{\sqrt{2\pi}}\left[-e^{-\frac{(c_{cd}-x_c)^2}{2\sigma^2}} + e^{-\frac{(c_{cd}-x_d)^2}{2\sigma^2}}\right]. \tag{3.9}
$$

Completely analogous to the optimization problem with respect to c_{ab}, whatever the values of x_a, x_b, x_c, x_d, the optimal choice for the hearer is to use criterion points $c_{bc} = (x_b + x_c)/2$ and $c_{cd} = (x_c + x_d)/2$.

We show in Figure 3.5 the results of the optimization for this case for a range of σ and k. Similar to the two-message case, with both decreasing cost parameter k and increasing but low noise level σ, the speaker uses more extreme signals to convey the same message. In contrast to the two-message case though, where the hearer adopted the same strategy for all values of k and σ, in the four-message case, the hearer must adjust the criterion points c_{ab} and c_{cd} in response to the change in the speaker's strategy. As the dashed lines in Figure 3.5 show, when the speaker uses more extreme signals, the hearer must compensate.

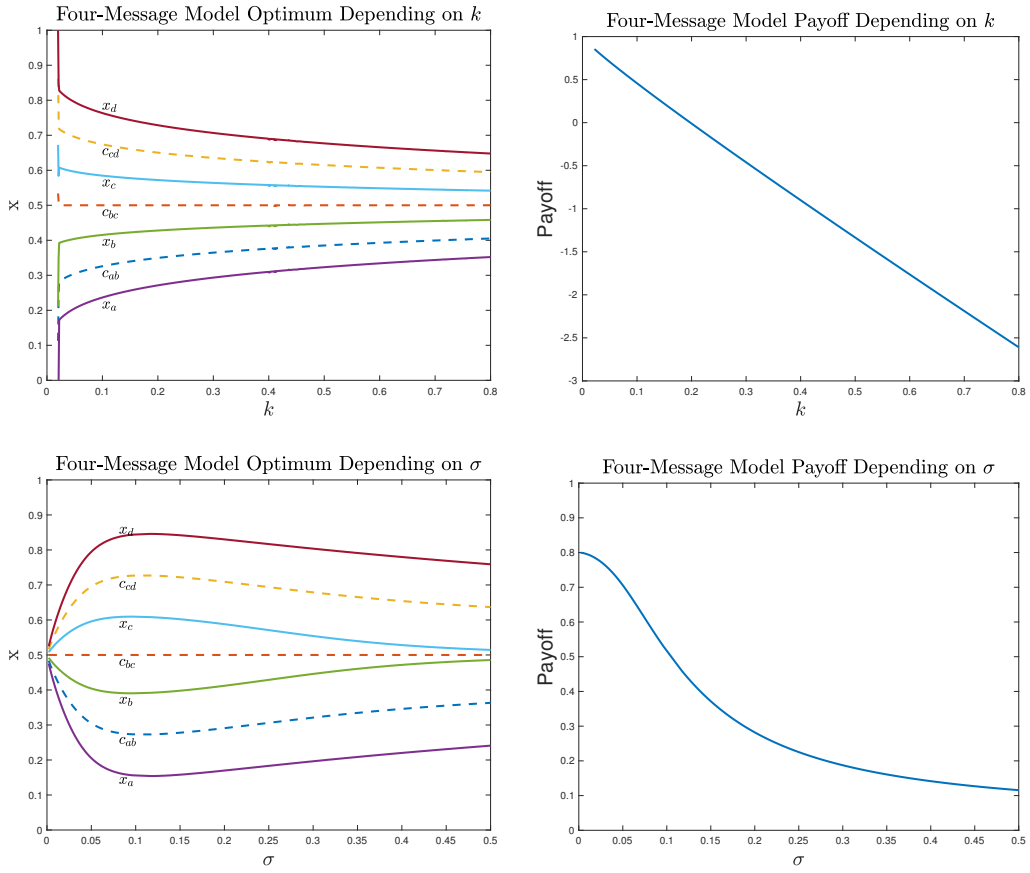

Figure 3.5: The four-signal model. *Top Left:* The value of the optimal x_a, x_b, x_c, x_d(*solidlines*) and c_{ab}, c_{bc}, c_{cd} (dashed lines) for varying k and $\sigma = 0.05$. *Top Right:* The value of the maximized payoff $E(x_a, x_b, x_c, x_d, c_{ab}, c_{bc}, c_{cd})$ for varying k and $\sigma = 0.05$. *Bottom Left:* The value of the optimal x_a, x_b, x_c, x_d (solid lines) and c_{ab}, c_{bc}, c_{cd} (dashed lines) for varying σ and $k = 0.05$. *Bottom Right:* The value of the maximized payoff $E(x_a, x_b, x_c, x_d, c_{ab}, c_{bc}, c_{cd})$ for varying σ and $k = 0.05$.

From the figures, we note that the speaker uses a larger portion of the phonetic space in the four-message game. This can be seen by comparing the range of x used for the signals

for any particular k and σ, as in Figures 2.5, 2.6 and 3.5. For example, when $k = 0.1$ and $\sigma = 0.05$, in the two-message case, the signals used range from about 0.4 to 0.6, whereas in the four-message case they range from about 0.25 to 0.75. This agrees with the Theory of Adaptive Dispersion [21], which postulates that the vowel space will become more dispersed when there are more phonemes needed to fit into the space. One way to explain the effect that the basic model in Chapter 2 has a smaller range of x than the four-message model is to use the fact that the smaller range already provides sufficient contrast between two messages and makes it possible for the receiver to recognize the message. When two more messages are added, and the number of contrasts that needs to be made rises to three, there is a need for expanding the phonetic space. But because more extreme signals x are more expensive, and signals less than 0 or greater than 1 are impossible, the range cannot be tripled. The net effect is that with more messages to transmit, the speaker expands the phonetic space used, while decreasing the spacing between the signals for distinct messages.

3.3 Conflicts Between Clearness and Comprehensibility

In the two-message model, how to categorize the received signal mainly relies on how the speaker maps the message to a continuous perceptual value, since the optimal strategy for the receiver is always the mid point between two sent signals. As we can see in the Chapter 2, the speaker needs to modify their strategy with the rise of noise in order to maximize payoff. But in this case the hearer does not need to make any adjustment to their strategy in response to the speaker's speech style, as shown by the flat dashed line labeled c in Figures 2.5, 2.6, which always keeps horizontal at $\frac{1}{2}$.

In the four-message model, we see from the dashed lines in Figure 3.5 that although the hearer only needs to select the mid points as the decision boundary, the optimal strategy for the hearer to distinguish the message pairs a and b, c and d is no longer a constant, since the two adjacent signals are not necessarily to be symmetric about a horizontal line. This implies that unlike the two-message model, the hearers in four-message model have to adjust their strategy as well when k or σ is changed, in order to optimally respond to the speaker's change of strategy.

Similarly, in [24] the strategy for the hearer to categorize the speech is context dependent. They interpret the signals according to the visual cues, like the speaker's gender and the judgements of speaker identity. In this section, we will discuss the situations where one of the players are not aware of the noise.

In previous chapters we assume that the noise condition in the game is completely clear to both speaker and hearer. There is another case where the information about the communication environment is asymmetric to the players, which is a common problem for our language users. In one case, the speaker was instructed to use clear speech, while the hearer is not aware that the signal was sent under the clear speech style so that the

hearer also does not know that the cue was exaggerated comparing to the plain one that the speaker intended to send. The signal is considered to be sent using the plain style, and will be interpreted using the criterion points for plain speech in the communication environment with low noise. Hence, the clear speech signal sent from high noise may be wrongly interpreted due to the incomplete information.

An example demonstrating this case well is the four-message model depending on σ shown in Figure 3.5 *bottom left*. Fix $k = 0.05$, and let there be two modes for the noise, where σ takes one of two values: $\sigma_1 = 0.01$ or $\sigma_2 = 0.05$. Suppose the speaker believes $\sigma = \sigma_2$ and intends to transmit message b and so utters a signal near $y = 0.4$. If the hearer thinks $\sigma = \sigma_1$, they will decode this as message a, which is the wrong message. The conflict between the classification and the real transmission is that a plain b under σ_1 has a similar signal to a clear a under σ_2, and so the hearer cannot tell them apart if they are not aware of the noise condition so that they do not know whether a plain or clear style is being used. Figure 3.6 demonstrates this case. If the receiver obtained a signal 0.4, the signal will be categorized to message a given that it was sent under the noise amplitude $\sigma = 0.01$, and message b for $\sigma = 0.05$.

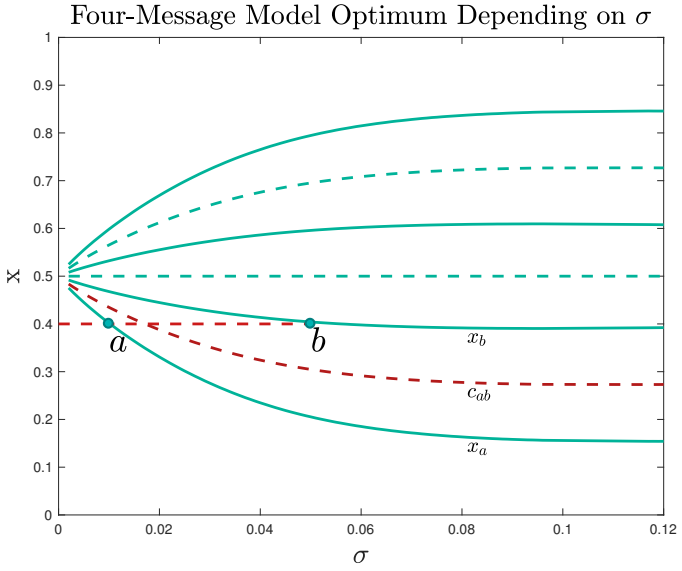

Figure 3.6: Optimum in four-message model depending on σ. When a signal with value 0.4 has been received, it would be classified as message a if the hearer thought the signal was sent under $\sigma = 0.01$ and message b if $\sigma = 0.05$.

For example, the difference between tense and lax vowels in English [20] indicates that if the listener does not know that the word was generated in clear speech, then the conflicts may occur. The paper [20] explored three English tense-lax vowel pairs, /i-ɪ/, /ɑ-ʌ/ and /u-ʊ/. The Fig 5 from [20] sketches the vowel space for the three pairs of tense-lax vowels in clear and plain speech style from the experiment data. Comparing to lax vowels (e.g., /ɪ/), tense vowel (e.g., /*i*/) are generated with more exaggerated articulatory movements,

longer target position, longer duration, and involved in a larger vowel space. In spoken English, though there are other important features to distinguish the tense and lax vowels [15], the differences between tense and lax vowel mentioned above have been shown to be perceptually significant. Imagine a speaker who wishes to be clearer in a lax vowel, then the speaker will lengthen the vowel and perform more exaggerated articulatory movements. If the listener does not know that the speaker was using a clear speech, there will be a great chance for the hearer to mistakenly categorize the received word as the corresponding tense counterparts. For a concrete example, a long "kid" might be difficult to distinguish from a plain "keyed". There is a conflict between a non-phonemic speech clarity effect (i.e. lengthening, more exaggerated gesture) and a phonemic contrast.

The similar phenomenon is also found in some animal communication systems. In [2], when the noise in the background environment is rising, both human and monkeys will increase the duration of their vocal. However, the improvement may not work for some situations. For example, [18] describes a small African rodent called Parotomys brantsii which uses duration of its call to represent the level of safety. When there are threats like a snake or human, long-duration calls will be given; while a relatively shorter call indicates the low-risk environment. We can conjecture, for these Brant's whistling rats, it is not useful to lengthen their calls in the presence of noise in order to make clearer, since extension of length has specific meanings.

To further study the case where clear speech may cause confusion to the listener, we assume that the speaker is aware of the noise amplitude σ, either 0.01 or 0.05, but the hearer does not know σ and is capable of using only one set of criterion points for both noise levels. What are the optional strategies for both players? One possibility is that the speaker insists upon a fixed uttering strategy though the noise is in one of two levels of amplitude. The second possibility is that the speaker thinks it is worth it to have two different sets of signals, one for each level of noise, though the hearer has no information about noise level. Next, our simulation will show that the predicted behaviour is a compromise between these two simpler strategies.

We assume that there are three different cases for the speaker's and hearer's awareness of the noise amplitude σ in our four-message model: both players share knowledge of noise, only hearer knows the noise level, and only receiver is aware of noise level. The speaker needs to communicate one of four equally-frequent possible messages, a, b, c, and d, each of which occur with probability $\frac{1}{4}$, and the effort parameter k for a signal is 0.05. What distinguishes this model from the four-message model in Section 3.2 is that in each trail of the game σ takes the value of either $\sigma_1 = 0.01$ or $\sigma_2 = 0.05$ with probability $\frac{1}{2}$, rather than a fixed noise level. Define the strategy for the speaker to transmit four message as $X = (x_a, x_b, x_c, x_d)$ and the strategy for the hearer to make decision based on the received signal as $C = (c_{ab}; c_{bc}; c_{cd})$. To explore the predictions of our model for these contexts, numerical optimization is computed shown as Figure 3.7. In the figure, we show the optimal

strategies for the speaker and hearer in the four-message, two noise-level game, in each of the three cases. The blue circles are signals x_a, x_b, x_c, x_d, and the red points c_{ab}, c_{bc}, c_{cd} are criterion points. To better emphasize the comparison between the cases, we add lines connecting corresponding signals and criterion points. The three cases and their optimal strategies are shown as the following:

- **Case i: Neither Oblivious.** The value of noise amplitude σ in each trail is known to both speaker and hearer. Since $X^{(1)}$ and $C^{(1)}$ only reply on σ_1, and $X^{(2)}$ and $C^{(2)}$ only reply on σ_2, thus the strategies in the two cases are independent. Hence both their strategies can be determined with respect to noise level. So we apply $X^{(1)}$ and $C^{(1)}$ for noise level σ_1 and $X^{(2)}$ and $C^{(2)}$ for noise level σ_2 as in the previous section. Equivalently, $X^{(1)}, X^{(2)}, C^{(1)}, C^{(2)}$ together maximize

  $$\frac{1}{2}E(X^{(1)}, C^{(1)}, k, \sigma_1) + \frac{1}{2}E(X^{(2)}, C^{(2)}, k, \sigma_2).$$

  Leftmost in the Figure 3.7, we show the strategies for each level of noise when both speaker and hearer are aware of the level of the noise. Since this case is simply the combination of two cases from the four-message model in Section 3.2 but with distinct noise levels, as we expect from the previous section, the speaker uses more extreme signal values when the noise is greater, and the hearer is able to take this into account in the setting of the criterion points. The payoff achieved in this case is 0.750.

- **Case ii: Oblivious Hearer.** The noise level σ is only known to the speaker. Speaker has strategies $X^{(1)}$ given the noise level as σ_1 and $X^{(2)}$ for σ_2, but the hearer only has C since the noise level is not known to the hearer. $X^{(1)}$, $X^{(2)}$ and C together maximize

  $$\frac{1}{2}E(X^{(1)}, C, k, \sigma_1) + \frac{1}{2}E(X^{(2)}, C, k, \sigma_2).$$

  In this case that only the speaker is aware of the level of the noise, the strategy for $\sigma_1 = 0.01$ and $\sigma_2 = 0.05$ are shown as the the middle of the Figure 3.7. When the noise level rise from 0.01 to 0.05, the speaker still emits more extreme signals, which is the same tendency as in the Neither Oblivious case. However, this effect is relatively more delicate: the range of the signals is greater than in the Neither Oblivious case when the noise level is σ_1, and the range of the signals is less than in the Neither Oblivious case when the noise level is σ_2. The fact that the noise level is not open to hearer means the speaker cannot deploy this strategy to full effect. The payoff achieved is now 0.741 which is lower than the optimal payoff in Neither Oblivious case, and hence a cost is paid for the hearer's ignorance.

- **Case iii: Both Oblivious.** Neither the speaker nor the hearer know the value of σ for each trail. On this case X and C are determined by maximizing

$$\frac{1}{2}F(X, C, k, \sigma_1) + \frac{1}{2}F(X, C, k, \sigma_2).$$

Rightmost in Figure 3.7, we show the strategy of two noise levels when neither speaker nor hearer is aware of the level in each trial. In this case where both players are oblivious, the strategies for the speaker and hearer are close to the ones in the second case (Oblivious Hearer) where $\sigma = 0.05$ and only the speaker knows the noise level. In case ii (Oblivious Hearer) and case iii (Both Oblivious), we can see that the receiver faces the same dilemma, the lack of information about noise level, and hence the receiver has similar strategy in the two cases. More interesting is that although there is a possible for noise amplitude to be 0.01 in some trials, the speaker in case iii still chooses the similar strategy to the one in case ii when $\sigma = 0.05$. But in case iii the additional exaggeration of the speaker in the trails where $\sigma_1 = 0.01$ is a kind of waste. This can be considered as that the players use a compromise of the strategies in the other cases, leading to a payoff of 0.729, worse than either of the other two cases.

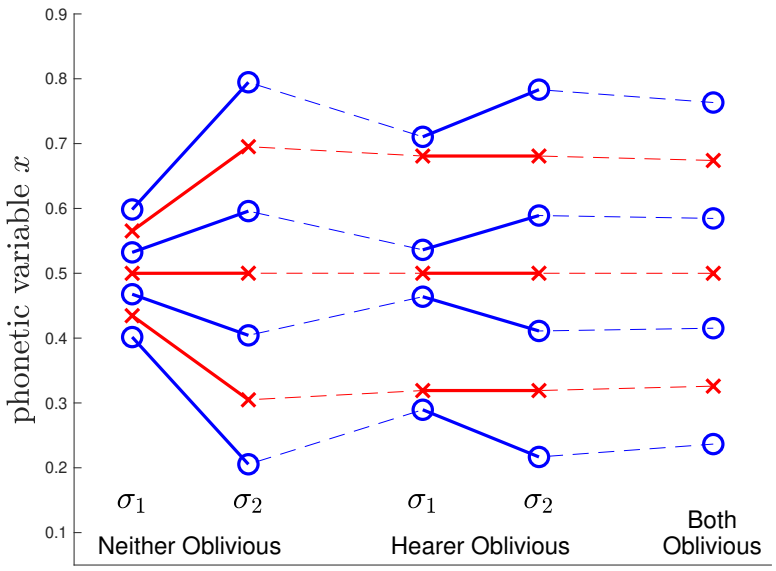

Figure 3.7: The optimal strategies for the speaker and the hearer in the four-signal model with two levels of noise. The condition is indicated along the x-axis. The y-axis indicates the value of the signal used for each of the four message by the speaker (in blue), and the three criterion points used by the hearer (in red).

Based on the results of this model, incomplete information of the hearer on whether a clear speech style is being used to mute the difference between plain and clear speech. We envision how to investigate this effect experimentally when the clear speech is needed. One way is to vary the instructions, sometimes explaining that the intended hearer will be aware

that the speech is clear, and sometimes not. The other way is to let both a speaker and hearer study together, with noise of different levels being played on separate headphones. Whether the speaker knows if the hearer has the same noise level or not can be manipulated, allowing this effect to be investigated.

Chapter 4

Multiple Phonetic Variables

In the previous chapters, we considered a communication game in which the speaker can vary only one phonetic variable in the signal. In real speech a multitude of different dimensions of a signal can be controlled. Our models can be expanded to handle more signal dimensions. This will allow us to model and study the effects of variables like amplitude (i.e. loudness or intensity) which are typically not used contrastively, as well as how speakers decide among multiple variables which to use as a contrastive one. In this chapter, we study the model with multiple phonetic variables.

4.1 Two Contrastive Features

Suppose we have two contrastive features x and y. We can imagine x and y as formants $F1$ and $F2$. The speaker has one of two messages a and b to send equally often. Similar to the basic model, the speaker's strategy is to select two points in the 2D vowel space: (x_1, y_1) and (x_2, y_2) for the signals a and b respectively. We assume that the original value selected by the speaker is perturbed by a standard multivariate Gaussian noise n such that $n \sim N([0,0]^T, I)$, where covariance matrix I is an identity matrix. The value that the hearer will receive is either $(x_1, y_1)^T + \sigma n$ and $(x_2, y_2)^T + \sigma n$, where σ is still a noise amplitude which contributes equally to both components in the vowel space. The hearer receives the values as

$$(x_1, y_1)^T + \sigma n \sim \mathcal{N}([x_1, y_1]^T, \sigma I) \text{ and } (x_2, y_2)^T + \sigma n \sim \mathcal{N}([x_2, y_2]^T, \sigma I). \tag{4.1}$$

The optimal choice for the hearer is to determine a curve in the 2D vowel space as the decision boundary, which sketched as Figure 4.1. The decision boundary, referred to as c, will separate the plane into two disjoint regions A and B. When the hearer receives signal y which located in the area A, he selects message a, otherwise he selects b. The probability

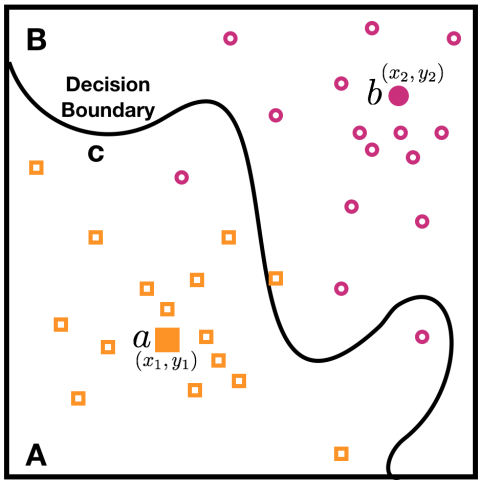

Figure 4.1: A schematic showing the communication game in two dimensions. The speaker is required to communicate one of two messages: a and b. They select signal values (x_1, y_1) and (x_2, y_2) which they transmit to the hearer. Noise in the communication channel leads to the hearer receiving a perturbed signal in the 2D vowel space which they classify as either a or b based on the decision boundary c.

that the receiver gets the correct message a can be written as

$$\begin{aligned}\mathbb{P}\left(\text{correct}|a\right) &= P\left([x_1, y_1]^T + \sigma n \in A\right) \\ &= \int\int_A f_a(x, y)\ dy\ dx \\ &= \int\int_A \frac{1}{2\pi} e^{-\frac{(x-x_1)^2+(y-y_1)^2}{2}}\ dy\ dx,\end{aligned} \tag{4.2}$$

where $(x_2, y_2)^T + \sigma n$ is the multivariate normal distribution with mean $\mu = (x_2, y_2)^T \in \mathbb{R}^2$ and covariance matrix $\Sigma = I = [1, 0; 0, 1] \in \mathbb{R}^{2\times 2}$, and its its probability density function $f(x, y)$ is given by

$$\begin{aligned}f(x, y) &= f\left(\vec{x} = [x, y]^T; \mu = [x_1, y_1]^T, \Sigma = [1, 0; 0, 1]\right) \\ &= \frac{1}{(2\pi)^{n/2}|\Sigma|^{n/2}} e^{-\frac{1}{2}(\vec{x}-\mu)^T\Sigma^{-1}(\vec{x}-\mu)} \\ &= \frac{1}{2\pi} e^{-\frac{(x-x_1)^2+(y-y_1)^2}{2}}.\end{aligned} \tag{4.3}$$

Similarly, we have the probability that the receiver gets the correct message b as

$$\begin{aligned}\mathbb{P}(\text{correct}|b) &= P\left([x_2, y_2]^T + \sigma n \in B\right) \\ &= \int\int_B \frac{1}{2\pi} e^{-\frac{(x-x_2)^2+(y-y_2)^2}{2}}\ dy\ dx,\end{aligned} \tag{4.4}$$

together giving us the probability of receiving the correct message as:

$$\begin{aligned} P(x_1, y_1, x_2, y_2, c) &= \mathbb{P}(\text{correct transmission}) \\ &= \mathbb{P}\left(\text{correct}|a\right)\mathbb{P}(a) + \mathbb{P}\left(\text{correct}|b\right)\mathbb{P}(b) \\ &= \frac{1}{2}\left(\int\int_A f_a(x,y)\ dy\ dx + \int\int_B f_b(x,y)\ dy\ dx\right), \end{aligned} \tag{4.5}$$

where $f_a(x,y) = \frac{1}{2\pi}e^{-\frac{(x-x_1)^2+(y-y_1)^2}{2}}$ and $f_b(x,y) = \frac{1}{2\pi}e^{-\frac{(x-x_2)^2+(y-y_2)^2}{2}}$.

We define the cost function for one signal (x, y) in the 2D model as

$$g(x,y) = \frac{x^2}{m^2} + \frac{y^2}{n^2}, \tag{4.6}$$

where $\frac{1}{m}$ represents the cost parameter for feature x and $\frac{1}{n}$ represents the cost parameter for feature y. Therefore, the payoff function for the 2D model can be expressed as

$$\begin{aligned} E(x_1, y_1, x_2, y_2, c) &= P(x_1, y_1, x_2, y_2, c) - (g(x_1, y_1) + g(x_2, y_2)) \\ &= \frac{1}{2}\left(\int\int_A f_a(x,y)\ dy\ dx + \int\int_B f_b(x,y)\ dy\ dx\right) \\ &- \frac{1}{2}\left(\frac{x_1^2}{m^2} + \frac{y_1^2}{n^2} + \frac{x_2^2}{m^2} + \frac{y_2^2}{n^2}\right). \end{aligned} \tag{4.7}$$

4.1.1 Decision Boundary is the Mid-perpendicular of Two Messages

We define the *mid-perpendicular* of two points A and B as the perpendicular bisector to segment AB as shown by the line m in Figure 4.2. In this subsection, we are going to prove that the optimal decision boundary in our 2D model is the mid-perpendicular of the two messages.

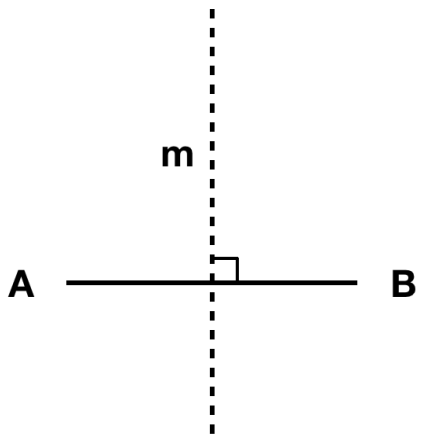

Figure 4.2: Mid-perpendicular of two points A and B

When the speaker's strategy is fixed, the cost of the messages is determined. In order to maximize the payoff function (4.7), the probability of correct transmission (4.5) should be maximized by the receiver's optimal strategy. As the optimal receiver's strategy will generate distinct regions A and B, if $f_a(x_0, y_0) > f_b(x_0, y_0)$ region A should contain the signal (x_0, y_0), otherwise the signal will be classified into region B. Therefore, the intersection of $f_a(x, y)$ and $f_b(x, y)$ on $x - y$ plane will be the optimal decision boundary for the receiver. The decision boundary $\left\{(x, y)|\frac{1}{2\pi}e^{-\frac{(x-x_1)^2+(y-y_1)^2}{2}} = \frac{1}{2\pi}e^{-\frac{(x-x_2)^2+(y-y_2)^2}{2}}\right\}$ can be proved to be the mid-perpendicular of the two messages (x_1, y_1) and (x_2, y_2), shown as Figure 4.3

To obtain the same conclusion, we can also consider the task of the receiver as the comparison of two probabilities, $P(a|(x, y))$ and $P(b|(x, y))$, where $P(a|(x, y))$ represents the probability of a given signal (x, y) be classified into message a, and similarly $P(b|(x, y))$ as probability of being message b.

Based on the Bayes Theorem, we have $P(a|(x, y))$ and $P(b|(x, y))$ as

$$P\left(a|(x, y)\right) = \frac{P\left((x, y)|a\right)P(a)}{P\left((x, y)\right)}, \tag{4.8}$$

$$P\left(b|(x, y)\right) = \frac{P\left((x, y)|b\right)P(b)}{P\left((x, y)\right)}, \tag{4.9}$$

where $P(a) = P(b) = \frac{1}{2}$, $P\left((x, y)|a\right) = f_a(x, y)$ and $P\left((x, y)|b\right) = f_b(x, y)$. Therefore, to get the optimal decision boundary, we only need to compare $f_a(x, y)$ and $f_b(x, y)$.

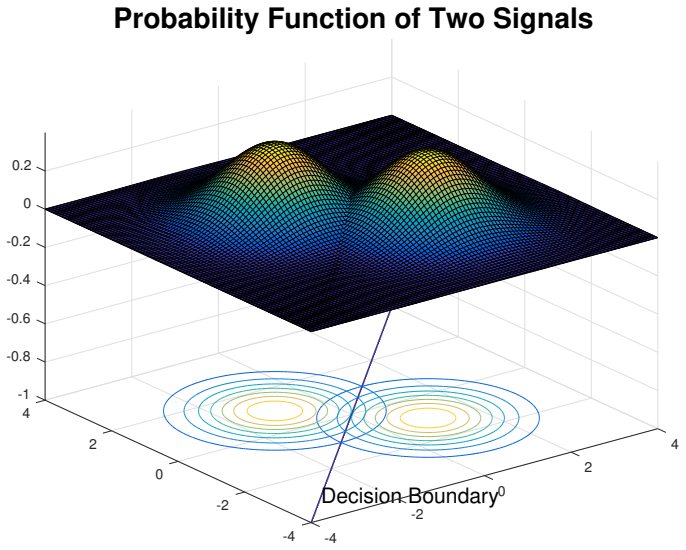

Figure 4.3: Probability density functions of two signals. Suppose the noise in the 2D model is a standard multivariate Gaussian random variable. The optimal decision boundary to decode the received signals is the perpendicular bisector to segment the speaker's strategies.

4.1.2 Analysis of the Optimum

Without loss of generality, we can assume that $y_1 \leq y_2$. First, suppose $y_1 \neq y_2$. Based on the speaker's strategy (x_1, y_1) for message a and (x_2, y_2) for message b, The decision boundary for hearer is the mid-perpendicular of (x_1, y_1) and (x_2, y_2):

$$y = -\frac{x_1 - x_2}{y_1 - y_2}(x - \frac{x_1 + x_2}{2}) + \frac{y_1 + y_2}{2}, \tag{4.10}$$

which we simply refer as $y = kx + b$.

Therefore, the payoff function (4.7) can be rewritten as

$$\begin{aligned} E &= \int_{-\infty}^{+\infty}\int_{-\infty}^{kx+b} \frac{1}{2\pi} e^{-\frac{(x-x_1)^2+(y-y_1)^2}{2\sigma^2}}\, dy\, dx + \int_{-\infty}^{+\infty}\int_{kx+b}^{+\infty} \frac{1}{2\pi} e^{-\frac{(x-x_2)^2+(y-y_2)^2}{2\sigma^2}}\, dy\, dx \\ &- \frac{1}{2}(\frac{x_1^2}{m^2} + \frac{y_1^2}{n^2}) - \frac{1}{2}(\frac{x_2^2}{m^2} + \frac{y_2^2}{n^2}) \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{a_1x+b_1} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}\, dy\, dx + \int_{-\infty}^{+\infty}\int_{a_2x+b_2}^{+\infty} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}\, dy\, dx \\ &- \frac{1}{2}(\frac{x_1^2}{m^2} + \frac{y_1^2}{n^2}) - \frac{1}{2}(\frac{x_2^2}{m^2} + \frac{y_2^2}{n^2}) \\ &= F(\frac{b_1}{\sqrt{1+a_1^2}}) + F(-\frac{b_2}{\sqrt{1+a_2^2}}) - \frac{1}{2}(\frac{x_1^2}{m^2} + \frac{y_1^2}{n^2}) - \frac{1}{2}(\frac{x_2^2}{m^2} + \frac{y_2^2}{n^2}) \end{aligned} \tag{4.11}$$

where a_1, b_1, a_2 and b_2 are

$$\begin{aligned} a_1 &= -\frac{x_1 - x_2}{y_1 - y_2}, \\ b_1 &= (\frac{x_1 - x_2}{y_1 - y_2} \times \frac{x_1 + x_2}{2} + \frac{y_1 + y_2}{2} - y_1 - \frac{x_1 - x_2}{y_1 - y_2} \times x_1)/\sigma, \\ a_2 &= -\frac{x_1 - x_2}{y_1 - y_2} = a_1, \\ b_2 &= (\frac{x_1 - x_2}{y_1 - y_2} \times \frac{x_1 + x_2}{2} + \frac{y_1 + y_2}{2} - y_2 - \frac{x_1 - x_2}{y_1 - y_2} \times x_2)/\sigma = -b_1. \end{aligned}$$

When $y_1 = y_2$, payoff function (4.11) can be expressed as the following:

$$\begin{aligned} E &= \int_{-\infty}^{+\infty}\int_{-\infty}^{x_1} \frac{1}{2\pi} e^{-\frac{(x-x_1)^2+(y-y_1)^2}{2\sigma^2}}\, dx\, dy + \int_{-\infty}^{+\infty}\int_{x_2}^{+\infty} \frac{1}{2\pi} e^{-\frac{(x-x_2)^2+(y-y_2)^2}{2\sigma^2}}\, dx\, dy \\ &- \frac{1}{2}(\frac{x_1^2}{m^2} + \frac{y_1^2}{n^2}) - \frac{1}{2}(\frac{x_2^2}{m^2} + \frac{y_2^2}{n^2}) \\ &= 2F(\frac{|x_2 - x_1|}{2\sigma}) - \frac{1}{2}(\frac{x_1^2}{m^2} + \frac{y_1^2}{n^2}) - \frac{1}{2}(\frac{x_2^2}{m^2} + \frac{y_2^2}{n^2}). \end{aligned} \tag{4.12}$$

Now we are going to find the optimal messages are symmetric by the origin about first showing that the probability of receiving the correct messages only relies on the distance

between the two messages, and then proving that with a fixed distance the symmetric messages will give the minimum cost.

- Given fixed distance between (x_1, y_1) and (x_2, y_2), the probability of receiving the correct message is a constant.

  Proof. When $y_1 < y_2$, based on the Function (4.11), the probability of receiving correct messages can be written as

$$\begin{aligned} P(x_1, y_1, x_2, y_2) &= F(\frac{b_1}{\sqrt{1+a_1^2}}) + F(-\frac{b_2}{\sqrt{1+a_2^2}}) \\ &= 2F(\frac{b_1}{\sqrt{1+a_1^2}}) \\ &= 2F(\frac{(\frac{x_1-x_2}{y_1-y_2} \times \frac{x_1+x_2}{2} + \frac{y_1+y_2}{2} - y_1 - \frac{x_1-x_2}{y_1-y_2} \times x_1)/\sigma}{\sqrt{1+(-\frac{x_1-x_2}{y_1-y_2})^2}}) \\ &= 2F(\frac{1}{2\sigma}\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}), \end{aligned} \tag{4.13}$$

  which only depends on the distance between (x_1, y_1) and (x_2, y_2).

  When $y_1 = y_2$, the probability is

$$\begin{aligned} P(x_1, y_1, x_2, y_2) &= 2F(\frac{|x_2 - x_1|}{2\sigma}) \\ &= 2F(\frac{1}{2\sigma}\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}), \end{aligned} \tag{4.14}$$

  which also only depends on the distance between (x_1, y_1) and (x_2, y_2). □

- With a fixed distance between (x_1, y_1) and (x_2, y_2), referred as $2d$, the symmetric messages about origin will give the minimum cost. More specific, when $m = n$, the optimal messages are symmetric about the origin on a circle; when $m < n$, the optimal messages are $(d, 0)$ and $(-d, 0)$.

  Proof. Assume that the distance between (x_1, y_1) and (x_2, y_2) is $2d$, the midpoint of (x_1, y_1) and (x_2, y_2) is (x_0, y_0), and the angle of the line connecting (x_1, y_1) and (x_2, y_2) with the horizontal line is α. Then (x_1, y_1) and (x_2, y_2) can be written as

$$\begin{aligned} (x_1, y_1) &= (x_0 - d \cdot \cos\alpha, y_0 - d \cdot \sin\alpha), \\ (x_2, y_2) &= (x_0 + d \cdot \cos\alpha, y_0 + d \cdot \sin\alpha). \end{aligned}$$

  The optimization problem aiming to minimize the cost function $C(x_1, y_1, x_2, y_2) = \frac{1}{2}(\frac{x_1^2}{m^2} + \frac{y_1^2}{n^2}) + \frac{1}{2}(\frac{x_2^2}{m^2} + \frac{y_2^2}{n^2})$, given the distance of (x_1, y_1) and (x_2, y_2), can be written

as the following:

$$\min \quad \frac{1}{2}\left(\frac{x_0^2 + d^2\cos^2\alpha}{m^2} + \frac{y_0^2 + d^2\sin^2\alpha}{n^2}\right). \tag{4.15}$$

Since (x_0, y_0), d and α are independent, moving (x_0, y_0) to the origin will minimize 4.15. We need to discuss the relation between m and n to further minimize $\frac{\cos^2\alpha}{m^2} + \frac{\sin^2\alpha}{n^2}$, which can be rewritten as

$$\frac{1}{m^2} + \sin^2\alpha(\frac{1}{n^2} - \frac{1}{m^2}) \tag{4.16}$$

If $m \neq n$, without loss of generality we can assume that $m < n$, to minimize 4.16, $\alpha = 0$. That is to say, $|x_1| = |x_2| = d$, $y_1 = y_2 = 0$. Similarly, if $m > n$, then $\alpha = \frac{\pi}{2}$ which means that the optimal locations for the messages will be $|y_1| = |y_2| = d$ and $x_1 = x_2 = 0$. When $m = n$, α can be any value, so the optimal (x_1, y_1) and (x_2, y_2) are on a circle and symmetric by the origin. □

From the above proof, we know that the optimum messages are symmetric about the origin. Suppose message a is (x, y) and then message b is $(-x, -y)$, based on the function 4.12 the payoff can be written as

$$E(x, y) = 2F(\frac{1}{\sigma}\sqrt{x^2 + y^2}) - (\frac{x^2}{m^2} + \frac{y^2}{n^2}). \tag{4.17}$$

We plot the payoff function and its vector field for both the case $m = n$ and $m \neq n$ in Figure 4.4. From the plot, we can also tell that when $m = n$, the best speaker's strategy sits on a circle. When $m \neq n$, for example $m > n$ in Figure 4.4 bottom, optimum message a can be $(x^*, 0)$ and message b can be $(-x^*, 0)$, which maximize the payoff function. Meanwhile, we notice that there are saddle points on the $y - axis$.

4.1.3 Numerical Optimization

Section 4.1.2 gives the analytical solutions for the cases where $m = n$ and $m \neq n$ respectively. In this section, we use MATLAB to calculate the numerical results for these two cases.

Equal Feature Effort Parameter $\frac{1}{n} = \frac{1}{m}$

Suppose $\frac{1}{n} = \frac{1}{m}$. As we can learn from Section 4.2.2, the optimum message would be symmetric about the origin and on a circle, as demonstrated in the Figure 4.5.

Fixing the value of m at 3, we increase the noise amplitude σ from 0 to 1. The radius of the circle will change with the increase of σ. Figure 4.6 *top left* shows how the optimal radius depends on σ for a fixed value of $m = n = 3$. As σ goes to 0, the radius goes to 0; as σ increases, the radius first increase to make more effort to raise the chance of receiving

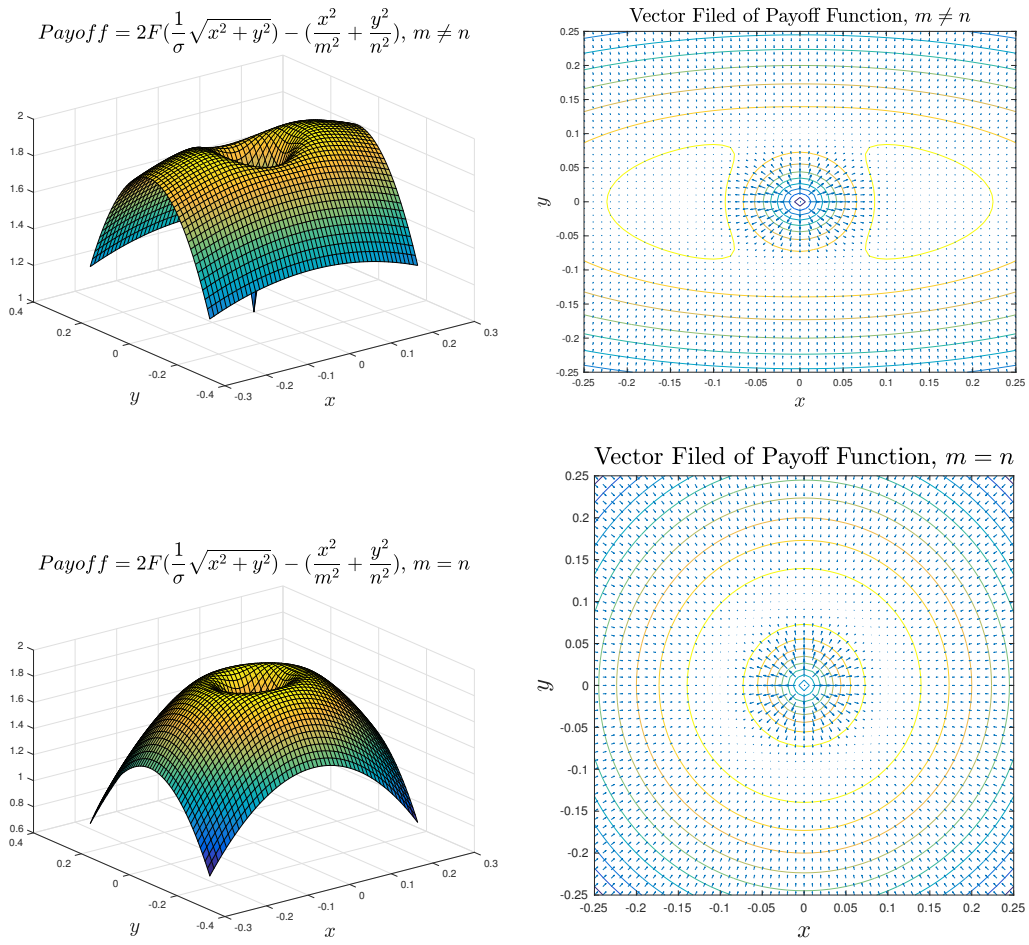

Figure 4.4: Payoff function in 2D constractive model. *Top Left:* The value of the payoff function given $m \neq n$. *Top Right:* The value of vector field of the payoff function given $m \neq n$. *Bottom Left:* The value of the payoff function given $m = n$. *Bottom Right:* The value of vector field of the payoff function given $m = n$.

correct signals, and then decrease after passing some certain threshold since it is too noisy to improve the communication. Figure 4.6 *top right* shows that as σ increases from 0 to 1, the optimized payoff keeps decreasing.

We fix the value of σ at 0.05 and vary m. Figure 4.6 *bottom left* shows how radius depend on $\frac{1}{m}$ for a fixed value of $\sigma = 0.05$. $\frac{1}{m}$ is equivalent to the effort parameter k in the 1D model, since smaller value of $\frac{1}{m}$ represents that it is less expensive to emit an exaggerated signal. For small values of $\frac{1}{m}$, as we might expect, the radius goes to infinity as $\frac{1}{m}$ goes to zero. When $\frac{1}{m}$ goes to infinity, the radius goes to 0, since the cost if emitting a signal becomes large compared to the benefit accurate communication. Figure 4.6 *bottom right* shows that as $\frac{1}{m}$ increases from 0 to 1, the optimized payoff keeps decreasing.

The equal feature effort parameter model is analogous to the previous 1D model.

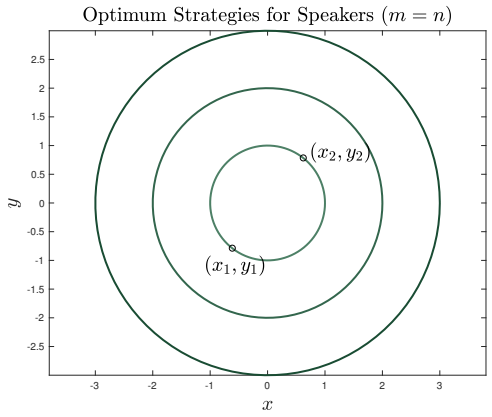

Figure 4.5: Optimum strategies for speaker in 2D contrastive model, given $m = n$. For a given communication environment (with noise level σ and effort parameter m and n fixed), the optimum strategies for the speaker would be two points on a circle which are symmetric about the origin. The radius of the circle is determined by the communication environment.

Unequal Feature Effort Parameter $\frac{1}{n} \neq \frac{1}{m}$

Suppose $\frac{1}{n} \neq \frac{1}{m}$. Without loss of generality, we can assume that $\frac{1}{n} < \frac{1}{m}$. From Section 4.2.2 we know that the more expensive features will be never used to distinguish the signals. Therefore, the speaker only makes effort with the cheaper feature. This case can be considered as the 1D case.

Figure 4.7 illustrates the optimum in the 2D contrastive model. The solid line represents the optimal positions of the speaker's strategy with varying m. When $m = n$, the two messages are located on any positions on the solid circle that are symmetric about the origin. Increasing m, optimal (x_1, y_1) and (x_2, y_2) keeps symmetry about the origin, but are constrained on the x-axis. As m increases, (x_1, y_1) moves towards the more negative direction on the $x - axis$ and conversely (x_2, y_2) moves towards the more positive direction on the $x - axis$, which makes sense since as m increases it is more and more cheaper to generate more exaggerated signals.

Fix the value of n as 2 and m as 3, the optimal strategy will be $(-x, 0)$ and $(x, 0)$. Figure 4.8 *top left* shows the optimal x with the change of σ from 0 to 1, and *top right* shows the optimized payoff is reducing as σ is growing. Figure 4.8 *bottom left* shows the optimal x with the change of m from 3 to $+\infty$ with $\sigma = 0.05$, and *bottom right* shows the optimized payoff is reducing as $\frac{1}{m}$ is growing. All these phenomenon are analogous to the 1D case.

4.2 Uncontrastive and Contrastive Features

According to [3], different languages have different acoustic cues to distinguish the words. The paper conducted experiments on three languages, English, Mandarin and Russian. The results reveals that the vowel quality is the strongest cue for all the three languages, and

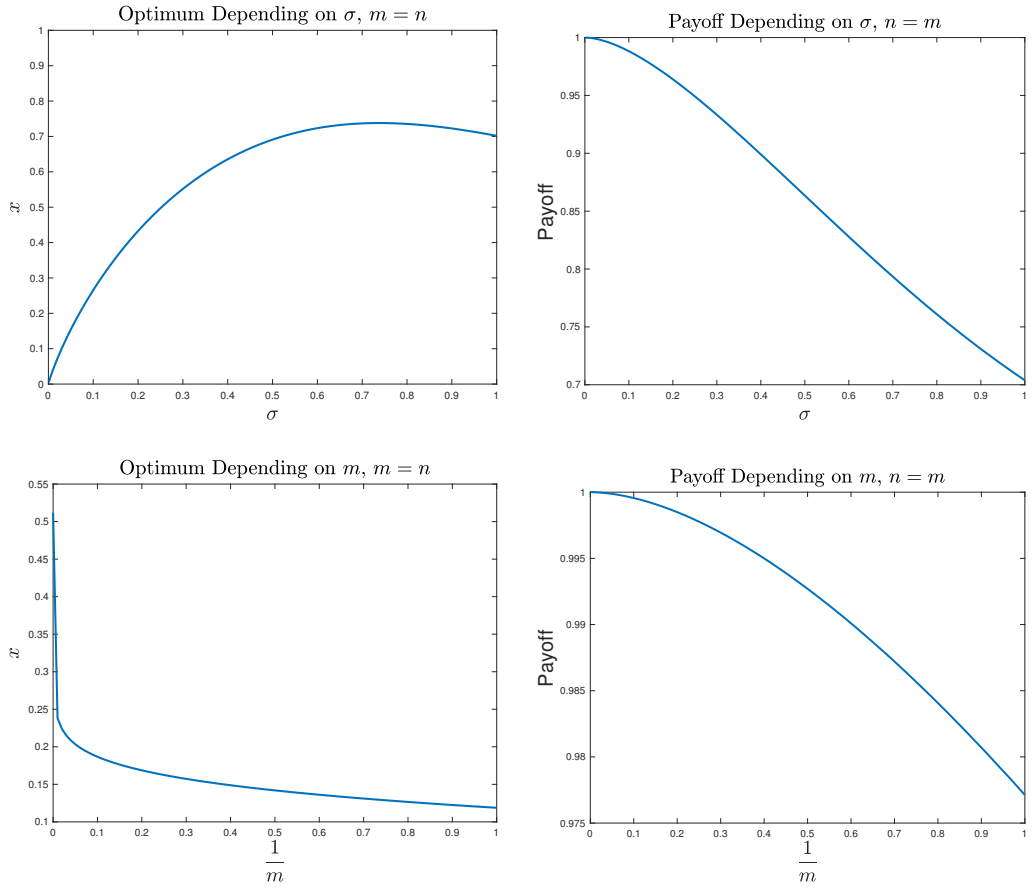

Figure 4.6: Numerical optimum of 2D constractive features model given $m = n$. *Top Left:* The value of the optimal radius for varying σ given $m = n = 3$. *Top Right:* The value of the maximized payoff for varying σ given $m = n = 3$. *Bottom Left:* The value of optimum radius for varying m given $\sigma = 0.05$. *Bottom Right:* The value of the maximized payoff for varying m given $\sigma = 0.05$.

pitch is the secondary cue for the Mandarin and English listeners. But pitch is not important for Russian listeners. In one language, among various phonetic cues, some cues, like formant, can be used to contrast different words when other cues, like duration, are not adequate for contrast. To compare tense-lax vowels in English, spectral difference is the primary cues and duration can serve as the secondary cue [20].

In one language, we can group its acoustic features into two: one kind of feature differs between different signals which is contrastive, and the other can not be used to compare different signals. We called the first kind of feature as contrastive feature and the second is called uncontrastive feature.

Contrastive features have been studied in Section 4.1. For uncontrastive feature, take volume as example, there are three properties for this kind of features:

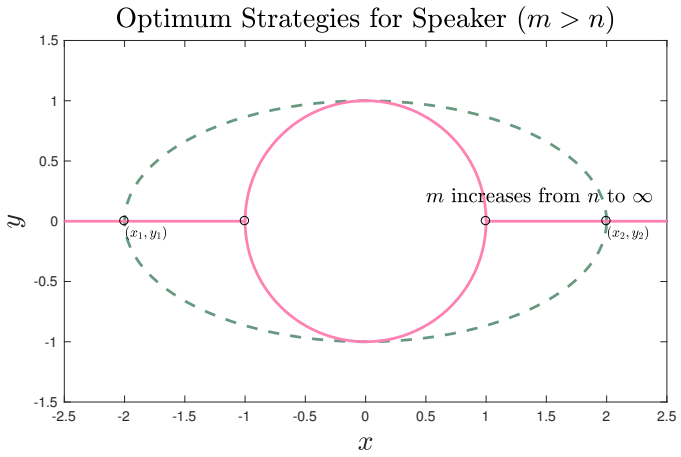

Figure 4.7: The optimal strategies for the speaker in 2D contrastive model given $m \geq n$. The solid line represents the optimal positions of the speaker's strategy varying m. When $m = n$, the two messages are located on any positions on the solid circle that are symmetric about the origin. Increasing m, optimal (x_1, y_1) and (x_2, y_2) move in the opposite directions on the x-axis to separate from each other.

- When an unconstrastive feature value is zero, the signal can not be recognized correctly. If the volume is set to zero, then the receiver can not hear the speaker which makes the communication fail.
- When noise is increasing, to make the signal better received, the value of the unconstrastive feature will increase. For example, people usually speaks more loudly in a noisy restaurant to have themselves more understandable. And the increase of the noise in the restaurant will make the original volume less useful.
- The effect of uncontrastive feature is independent among different signals. For example, varying the volume will not affect the contribution of duration in distinguishing different signals.

We combine these ideas by defining a probability function to receive correct signal according to uncontrastive feature x and the amplitude of noise σ as

$$\tilde{P}(x, \sigma) = 1 - e^{-\frac{x}{\sigma}},$$

which is sketched in Figure 4.9 with multiple values of σ.

We chose this form for $\tilde{P}$ because it means that

- It ranges between 0 and 1 and is positive, which can serve as a distribution function;
- As the feature value x increases, $\tilde{P}$ increases and goes to 1, satisfying the property (2) of the uncontrastive features;

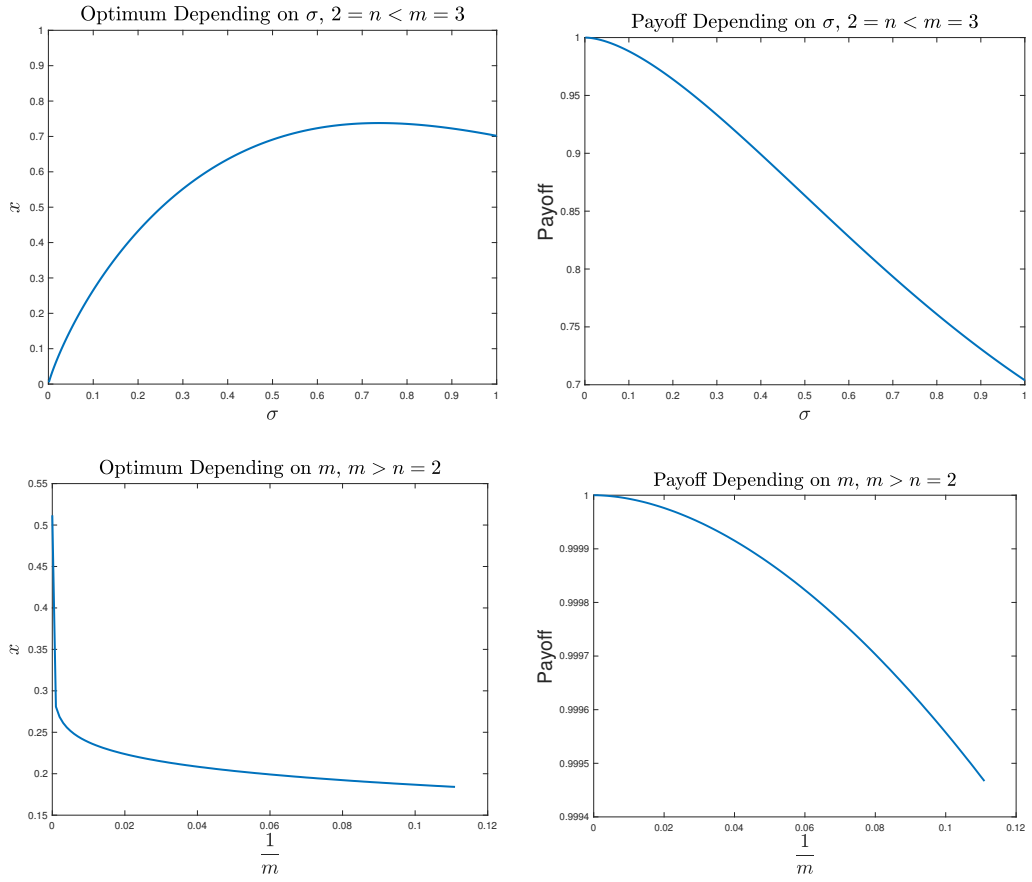

Figure 4.8: Numerical optimum of 2D constractive features model given $n < m$. *Top Left:* The value of the optimal x for varying σ given $m = 3$ and $n = 2$. *Top Right:* The value of the maximized payoff for varying σ given $m = 3$ and $n = 2$. *Bottom Left:* The value of optimum x for varying m from 3 to $+\infty$ given $\sigma = 0.05$. *Bottom Right:* The value of the maximized payoff for varying m from 3 to $+\infty$ given $\sigma = 0.05$.

- As the feature value x is approaching 0, $\tilde{P}$ goes to 0, satisfying the property (1);
- Most of all, for fixed uncontrastive feature x, with the increase of the noise amplitude σ the probability $\tilde{P}$ will decrease.

Now, we are going to study the game theoretic model with contrastive and uncontrastive features. Suppose the contrastive feature is marked as y and standard Gaussian distributed noise n will be added to the received the contrastive feature, which will be $y + n\sigma$. Suppose the uncontrastive feature is component x. Suppose there are two messages in our game. The speaker's task is the same as before, but now it will call the two signals (x_1, y_1) and (x_2, y_2) where x is uncontrastive feature and y is contrastive feature. The receiver needs to select a value c to distinguish the received value of the contrastive feature.

In the 2D contrastive and uncontrastive model, we assume that the uncontrastive feature takes the role of helping contrastive feature to more accurately transmit signals. The total

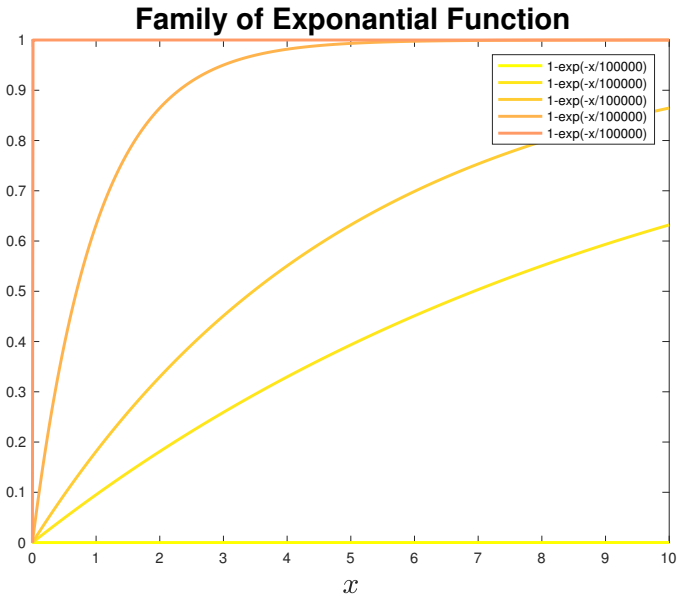

Figure 4.9: $\tilde{P} = 1 - e^{\frac{x}{\sigma}}$, the function describing the probability of uncontrastive feature.

probability to get correct signal based on both contrastive feature and uncontrastive feature can be the multiplication of the two probabilities. Based on the definition of probability corresponding to the contrastive and uncontrastive feature in Function (2.3) and (4.18), the probability to receive correct signals a and b can be defined as the product of the two probabilities:

$$\begin{aligned} P(\text{correct}|a) &= P(y_1 + n\sigma \leq c)\tilde{P}(x_1, \sigma) \\ &= F\left(\frac{c - y_1}{\sigma}\right)\left(1 - e^{-\frac{x_1}{\sigma}}\right), \end{aligned} \tag{4.18}$$

$$\begin{aligned} P(\text{correct}|b) &= P(y_2 + n\sigma > c)\tilde{P}(x_2, \sigma) \\ &= F\left(-\frac{c - y_2}{\sigma}\right)\left(1 - e^{-\frac{x_2}{\sigma}}\right), \end{aligned} \tag{4.19}$$

We assume that the contrastive feature y is more expensive than the uncontrastive feature x. Therefore, for the signal (x, y), the cost function can be defined as

$$C(x, y) = \frac{x^2}{2} + y^2. \tag{4.20}$$

The payoff function can be written as

$$
\begin{aligned}
E &= P(\text{correct}|a) + P(\text{correct}|b) - C(x_1, y_1) - C(x_2, y_2) \\
&= F\left(\frac{c - y_1}{\sigma}\right)\left(1 - e^{-\frac{x_1}{\sigma}}\right) + F\left(-\frac{c - y_2}{\sigma}\right)\left(1 - e^{-\frac{x_2}{\sigma}}\right) - \left[\left(\frac{x_1^2}{2} + y_1^2\right) + \left(\frac{x_2^2}{2} + y_2^2\right)\right]
\end{aligned}
\tag{4.21}
$$

Solve the optimization problem (4.21) using MATLAB with the change of noise amplitude σ. We can see that Figure 4.10 shows the numerical results of the 2D contrastive and uncontrastive features. As noise amplitude σ increases, uncontrastive feature $x_1 = x_2$ are increasing, and meanwhile the two signals become more and more distinct in contrastive feature y_1 and y_2. After a certain value of σ, both the uncontrastive and contrastive features begin to decrease. Figure 4.10 *right* shows that the optimal payoff is decreasing with the augment of noise.

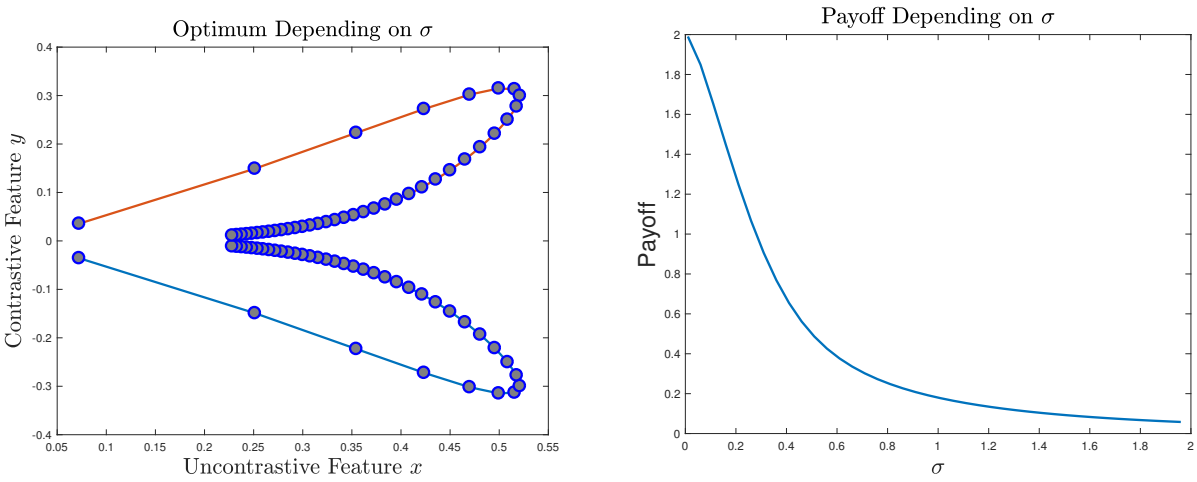

Figure 4.10: Optimum strategy in 2D model depending on σ. *Left:* Optimum for 2D model with uncontrastive and contrastive feature. *Right:* Payoff depending on σ.

Chapter 5

Conclusions

In this thesis, we have developed a method to model the communication between two players in adverse conditions using game theoretic framework. In the basic model with one phonetic variable and two messages needed to be sent, we show how the optimum strategies of speaker and receiver vary with the change of noise and the cost of effort respectively. When noise increases, the speaker will distinguish the two messages in order to make the receiver be able to recognize the correct one. After a certain noise level, the speaker gives up make efforts to distinguish the messages since the noise is large enough and it is not worthy to make more effort. When it is more and more difficult to make effort, the sender will try less harder to emit signals. These observations agree with the Lindblom's H&H theory [22] that the speaker has to strike a balance between the effort expended in transmitting the messages and the probability that the receiver can get a correct message.

Moreover, we study how the a priori probability of the potential messages needed to be sent affects the strategies of the players and the payoff in the game. We found that the speaker will expend more effort on the message which is more frequently sent. The payoff in the game with unequally likely sent messages is greater than the payoff in the game with equally likely messages. When we further expand our two-message model to four-message, the speaker's strategies change depending on the noise level and effort parameter in the similar way as the two-message model. The speaker in four-message game consumes more phonetic space than the two-message model, which is in accordance with the Adaptive Dispersion Theory [21].

We then consider a communication model which handles two phonetic variables. When it is equally diffifcult to make effort on both of the two phonetic features, the messages will locate on a circle and be symmetric about the center. If it is easier to exert energy on one feature, then the speaker will not make any effort on the more expensive feature.

Finally, we define contrastive features as the ones which can be used to distinguish different messages, and the uncontrastive features can assist contrastive features to distinguish signals but can not distinguish messages barely on themselves. When the noise is increasing, the uncontrastive features of the two signals are identical and both are increasing, while

the contrastive features of the two signals are keeping away from each other. Similarly to the 1D case, when the noise passes a certain value, the two features of each signal begin to decrease.

Bibliography

[1] Anton Benz, Gerhard Jäger, and Robert Van Rooij. An introduction to game theory for linguists. In *Game theory and pragmatics*, pages 1–82. Springer, 2006.

[2] Henrik Brumm, Rouven Schmidt, and Lars Schrader. Noise-dependent vocal plasticity in domestic fowl. *Animal behaviour*, 78(3):741–746, 2009.

[3] Anna Chrabaszcz, Matthew Winn, Candise Y Lin, and William J Idsardi. Acoustic cues to perception of word stress by english, mandarin, and russian speakers. *Journal of speech, language, and hearing research*, 57(4):1468–1479, 2014.

[4] Alan Cruttenden. *Gimson's pronunciation of English*. Routledge, 2013.

[5] Mike Davenport and Stephen J Hannahs. *Introducing phonetics and phonology*. Routledge, 2013.

[6] Gerald Patrick Delahunty and James J Garvey. *The English language: from sound to sense*. WAC Clearinghouse, 2010.

[7] Matthew Jared DeVos and Deborah A Kent. *Game Theory: A Playful Introduction*. American Mathematical Society, 2016.

[8] Sarah Hargus Ferguson and Diane Kewley-Port. Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 112(1):259–271, 2002.

[9] Takako Fujiwara-Greve. *Non-cooperative game theory*, volume 1. Springer, 2015.

[10] Sean A Fulop. *Speech spectrum analysis*. Springer Science & Business Media, 2011.

[11] Robert P Gilles. *The cooperative game theory of networks and hierarchies*, volume 44. Springer Science & Business Media, 2010.

[12] Gerhard Jäger. Applications of game theory in linguistics. *Language and Linguistics compass*, 2(3):406–421, 2008.

[13] Gerhard Jäger. Evolutionary stability conditions for signaling games with costly signals. *Journal of theoretical biology*, 253(1):131–141, 2008.

[14] Gerhard Jäger, Lars P Metzger, and Frank Riedel. Voronoi languages: Equilibria in cheap-talk games with high-dimensional types and few signals. *Games and economic behavior*, 73(2):517–537, 2011.

[15] Dennis H Klatt. Linguistic uses of segmental duration in english: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, 59(5):1208–1221, 1976.

[16] Peter Ladefoged and Sandra Ferrari Disner. *Vowels and consonants*. John Wiley & Sons, 2012.

[17] Jennifer Lam, Kris Tjaden, and Greg Wilding. Acoustics of clear speech: Effect of instruction. *Journal of Speech, Language, and Hearing Research*, 55(6):1807–1821, 2012.

[18] Aliza Le Roux, Tim P Jackson, and Michael I Cherry. Does brants'whistling rat (parotomys brantsii) use an urgency-based alarm system in reaction to aerial and terrestrial predators? *Behaviour*, 138(6):757–773, 2001.

[19] Jaewook Lee, Hussnain Ali, Ali Ziaei, Emily A Tobey, and John HL Hansen. The lombard effect observed in speech produced by cochlear implant users in noisy environments: A naturalistic study. *The Journal of the Acoustical Society of America*, 141(4):2788–2799, 2017.

[20] Keith KW Leung, Allard Jongman, Yue Wang, and Joan A Sereno. Acoustic characteristics of clearly spoken english tense and lax vowels. *The Journal of the Acoustical Society of America*, 140(1):45–58, 2016.

[21] Johan Liljencrants and Björn Lindblom. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, pages 839–862, 1972.

[22] Björn Lindblom. Explaining phonetic variation: A sketch of the h&h theory. In *Speech production and speech modelling*, pages 403–439. Springer, 1990.

[23] Neil A Macmillan. Signal detection theory. *Stevens' handbook of experimental psychology*, 4:43–90, 2002.

[24] Bob McMurray and Allard Jongman. What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, 118(2):219, 2011.

[25] Michael Mesterton-Gibbons and Mike Mesterton-Gibbons. *An introduction to game-theoretic modelling*. Number 11. American Mathematical Soc., 2001.

[26] Oskar Morgenstern and John Von Neumann. *Theory of games and economic behavior*. Princeton university press, 1953.

[27] John Nash. Non-cooperative games. *Annals of mathematics*, pages 286–295, 1951.

[28] J Desmond O'Connor and John LM Trim. Vowel, consonant, and syllable? A phonological definition. *Word*, 9(2):103–122, 1953.

[29] W Van Summers, David B Pisoni, Robert H Bernacki, Robert I Pedlow, and Michael A Stokes. Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America*, 84(3):917–928, 1988.

[30] Wilson P Tanner Jr and John A Swets. A decision-making theory of visual detection. *Psychological review*, 61(6):401, 1954.

[31] Vivien C Tartter, Hilary Gomes, and Elissa Litwin. Some acoustic effects of listening to noise on speech production. *The Journal of the Acoustical Society of America*, 94(4):2437–2440, 1993.

[32] Paul Tupper, Jie Jian, Keith Leung, and Yue Wang. Game theoretic models of clear versus plain speech. *CogSci 2018 Proceedings*.

[33] Jörgen W Weibull. *Evolutionary game theory*. MIT press, 1997.