

Title	雑音環境下における聴覚フィードバックが母音発話に与える影響に関する研究
Author(s)	西垣, 朋哉
Citation	
Issue Date	2020-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/16706
Rights	
Description	Supervisor: 赤木 正人, 先端科学技術研究科, 修士 (情報科学)

修士論文

雑音環境下における聴覚フィードバックが母音発話
に与える影響に関する研究

1810138 西垣 朋哉

主指導教員 赤木 正人
審査委員主査 赤木 正人
審査委員 鷗木 祐史
党 建武
吉高 淳夫

北陸先端科学技術大学院大学
先端科学技術研究科
(情報科学)

令和2年3月

Abstract

Our living environments are full of various noises. In such noisy conditions, humans involuntarily change their way of speaking for intelligible speech. This phenomenon is called “Lombard effect”. The uttered speech due to the Lombard effect also is called “Lombard speech”. In the Lombard speech, speech intensity, spectral tilt, formant frequency, fundamental frequency (F0), and duration or speaking rate are different from neutrally spoken speech. Moreover, the Lombard effect occurs not only in human but also in many other animals such as primates, birds, cats, whales, bats, and frogs. With regard to formants, some previous studies found that frequency of the first formant (F1) systematically increases by Lombard effect. Some studies reported that frequency of the second formant (F2) also increases, but not systematical as that of F1. Such variations in acoustic characteristics by the Lombard effect are dependent on tongue, jaw, and lip movements. The tongue position of vowels in Lombard effect is on average lower than that during neutral speech. Garnier et al. also found correlation of tongue and lip movements not only with F1 but also with F2. Stowe and Golob reported that speech intensity, duration, and F0 of uttered speech increase in bandlimited broadband noise (0.5–4 kHz band). However, with notched noise (0.5-4 kHz notch), their acoustic features did not change. These results indicated that characteristics of Lombard effect depend on properties of the noises. Matsumoto and Akagii investigated what strategies speakers use to utter intelligible speech under various noisy conditions. However, this research did not grasp what strategies speakers use to make uttered speech more intelligible in the noisy conditions. Therefore, this paper aims to grasp the tendency what strategies speakers make uttered speech more intelligible under various noisy conditions.

In this paper, seven types of noise were used, low-pass noise (LPN), high-pass noise (HPN), band-pass noise (BPN), notched noise (NN), and pink noise (PN). In order to verify whether the experimental environment is valid and whether the Lombard effect occurs, we investigated variations of speech when uttering in the same noise (PN) as in the previous study. PN is generated by applying a high-band attenuation filter to broadband white noise. Moreover, in PN, low frequency component are emphasized more than those in the other noises. The amplitude is calibrated to broadband noise (1–22,050 Hz) generated so that the RMS value is 70 dB and 80 dB. The noises used in this paper are adjusted to have the same dB/Hz as the broadband noise.

In order to analyze the frequencies and amplitudes of F1 and F2, four adult speakers (2 males and 2 females) age 23 to 24 participated in the recording. They have no obstacle to hearing function. The vowel utterances of each speaker were

recorded in quiet conditions in advance. While this recording, the speakers were asked to wear an open-air type headphone (STAX SR-L500). In this time, the noises were not presented from the headphone. Five speech data were obtained for each type of the vowels. The sampling frequency was 44,100 Hz.

The noises for this experiment were generated to correspond to F1 and F2 frequencies of each vowel and speaker in order to compare Lombard speech with neutral speech. Therefore, formant frequencies of recording speech were calculated by Acoustic core. This is a speech analysis software based on LPC. Then, this paper calculated the estimated frequencies of F1, F2 and F3 of the vowels of each speaker. In order to consider the effect of generated noises for the formants on auditory perception, this paper also calculated the excitation patterns based on the previous research.

Based on the result of formant frequency analysis, the frequency bands of the presentation noises are determined so as to correspond to the formant frequency and vowel types of each speaker. Cut-off frequencies were set to the mid frequencies of F1-F2 and F2-F3 for each speaker and vowel. These values depend on the types of noise.

To investigate variations of the frequencies and amplitudes of formants when the speakers utter the vowels under the noisy conditions, the noises were presented from the headphone. The noises were presented for each vowel in the order of BPN (Cut-off: mid frequency of F1-F2 to mid frequency of F2-F3), LPN_F12 (Cut-off: mid frequency of F1-F2), LPN_F23 (Cut-off: mid frequency of F2-F3), HPN_F12 (Cut-off: mid frequency of F1-F2), HPN_F23 (Cut-off: mid frequency of F2-F3), NN (Cut-off: mid frequency of F1-F2 to mid frequency of F2-F3) and PN. The noises also were presented in the order of 70 dB and 80 dB. For each vowel, 14 tasks (7 noises x 2 intensities) were prepared and a total of 70 tasks were carried out. One task is to utter the same vowel five times within 20 seconds under the noisy conditions. From this experiment, 350 speech data were obtained from each subject.

According to the results, this study found the tendency that regardless of the types of noise, F1 changes its frequency and F2 changes its amplitude mainly according to sound pressure level of the noises. These variations depend on types of noise. In detail, it was found that the greater the influence on F1 and F2 on the excitation patterns is, the larger the variations for utterance are. These results can be explained using excitation patterns of noise. The effect of PN is larger than that of other noise for utterance. PN is more emphasized in low frequency than the other noises. On the other hand, from the result of HPN_F23, the variations are smaller than that of the other noises. This noise has smaller the influence on F2 and F1 than the other noises. Therefore, from these results, it is suggested that

the noises in the low frequency influence on F1 and F2 more than the others when uttering vowels. Moreover, it is suggested that the positions of the tongue become lower as the sound pressure level of the noises increases. On the other hand, the horizontal positions of the tongue have no tendency as the sound pressure level of the noises increases. This was pointed out in previous study. Increase in F2 amplitude seems to be related to the glottal-fold source signal characteristics.

目次

第1章	序論	1
1.1	研究背景	1
1.2	先行研究	3
1.3	研究目的	3
1.4	本論文の構成	3
第2章	研究の着眼点	6
2.1	音声生成	6
2.2	音声知覚	8
2.3	聴覚フィードバック	10
2.4	研究計画	12
第3章	雑音環境下における母音発話実験手法	13
3.1	概要	13
3.2	実験刺激	13
3.3	静かな環境での発話の収録と分析	19
第4章	雑音環境下における母音発話の変化	22
4.1	フォルマント周波数に対応した刺激の生成	22
4.2	雑音環境下における発話の収録	22
4.3	雑音環境下におけるフォルマントの変化	25
4.3.1	結果	25
4.3.2	考察	25
4.4	雑音環境における興奮パターン	42
4.4.1	結果	42
4.4.2	考察	42
4.5	刺激が発話に与える影響とそれらの変化	50
4.5.1	結果	50
4.5.2	考察	50
第5章	総合考察	55
5.1	雑音による発話の変化	55
5.2	背景雑音の音圧レベルによる発話の変化	55

5.3 背景雑音の種類による発話の変化	55
第6章 結論	57
6.1 明らかにしたこと	57
6.2 のこされた課題	57
謝辞	58
研究業績	59

目 次

1.1	雑音環境下での発話	2
1.2	本論文の構成	5
2.1	子供, 若者, 大人が発話した日本語5母音のF1・F2の分布 (参照 [25])	7
2.2	1 kHzの純音を20 dBから80 dBまで20 dBずつ付加した際の興奮パターン	9
2.3	発話運動メカニズムの概念 (参照 [1])	11
3.1	実験刺激 (PN)	15
3.2	実験刺激 (BPN, NN)	16
3.3	実験刺激 (LPN_F12, LPN_F23)	17
3.4	実験刺激 (HPN_F12, HPN_F23)	18
3.5	実験環境	20
4.1	音声フィードバック情報に70 dBのBPNを付加した際のスペクトル (被験者1, /a/)	23
4.2	音声フィードバック情報に70 dBのBPNを付加した際の興奮パターン (被験者1, /a/)	24
4.3	背景雑音としてピンク雑音を付加した際の被験者毎のフォルマント周波数の変化	27
4.4	背景雑音としてピンク雑音を付加した際の被験者毎のフォルマント振幅の変化	28
4.5	背景雑音としてLPN_F12を付加した際の被験者毎のフォルマント周波数の変化	29
4.6	背景雑音としてLPN_F12を付加した際の被験者毎のフォルマント振幅の変化	30
4.7	背景雑音としてLPN_F23を付加した際の被験者毎のフォルマント周波数の変化	31
4.8	背景雑音としてLPN_F23を付加した際の被験者毎のフォルマント振幅の変化	32
4.9	背景雑音としてHPN_F12を付加した際の被験者毎のフォルマント周波数の変化	33

4.10	背景雑音として HPN_F12 を付加した際の被験者毎のフォルマント 振幅の変化	34
4.11	背景雑音として HPN_F23 を付加した際の被験者毎のフォルマント 周波数の変化	35
4.12	背景雑音として HPN_F23 を付加した際の被験者毎のフォルマント 振幅の変化	36
4.13	背景雑音として NN を付加した際の被験者毎のフォルマント周波数 の変化	37
4.14	背景雑音として NN を付加した際の被験者毎のフォルマント振幅の 変化	38
4.15	背景雑音として BPN を付加した際の被験者毎のフォルマント周波 数の変化	39
4.16	背景雑音として BPN を付加した際の被験者毎のフォルマント振幅 の変化	40
4.17	雑音毎の音圧レベルによる F1 と F2 の周波数・振幅の変化	41
4.18	背景雑音として PN を付加した際の興奮パターン（被験者 1,/a/)	43
4.19	背景雑音として HPN_F12 を付加した際の興奮パターン（被験者 1,/a/)	44
4.20	背景雑音として HPN_F23 を付加した際の興奮パターン（被験者 1,/a/)	45
4.21	背景雑音として LPN_F12 を付加した際の興奮パターン（被験者 1,/a/) 46	
4.22	背景雑音として LPN_F23 を付加した際の興奮パターン（被験者 1,/a/) 47	
4.23	背景雑音として BPN を付加した際の興奮パターン（被験者 1,/a/)	48
4.24	背景雑音として NN を付加した際の興奮パターン（被験者 1,/a/)	49
4.25	刺激が F1 に与える影響と F1 の周波数の変化	51
4.26	刺激が F2 に与える影響と F2 の周波数の変化	52
4.27	刺激が F1 に与える影響と F1 の振幅の変化	53
4.28	刺激が F2 に与える影響と F2 の振幅の変化	54

表 目 次

3.1	被験者 1 における平静発話時の母音のフォルマント周波数（成人男性）	21
3.2	被験者 2 における平静発話時の母音のフォルマント周波数（成人男性）	21
3.3	被験者 3 における平静発話時の母音のフォルマント周波数（成人女性）	21
3.4	被験者 4 における平静発話時の母音のフォルマント周波数（成人女性）	21

第1章 序論

1.1 研究背景

音声によってコミュニケーションを行う。これは、ヒトが意思疎通するにあたって最も重要な手段である。ヒトは、音声発話過程とその逆過程である音声知覚の繰り返しにより、言語音声知覚機構の発達を遂げてきた。つまり、音声コミュニケーションを円滑に保つためには、発話と知覚の双方が一体となって働く必要がある。そのため、発話者が音声を発する際には、自分自身の耳で音声を聞きとりながら、その強さ、高さを調整することで聞き手と良好な音声コミュニケーションを楽しむことが出来る。この機能を聴覚フィードバックと呼び、これによって話者が発話器官を制御することで、聞き手に正確に音声を伝達することが出来る [1]。この聴覚フィードバックは、発話において重要な役割を演じている。これは、遅延聴覚フィードバック (DAF) 実験から明らかである。この実験は、話者の発話した音声を遅延して話者にフィードバックするというものである。この実験では、話者に吃音や発話速度の低下といった現象が観測された [2, 3]。

我々の日常生活には、エアコンの稼働音、自動車の走行音、飲食店のBGMなどといった様々な外界ノイズが存在する。このような外界ノイズは、当然音声コミュニケーションを行う際には障害となる。そのような環境下でも良好なコミュニケーションを行うためには、聞き手への了解性を高めるため、発話者は音声の明瞭性を向上させる必要がある。この時、発話者は聴覚フィードバックによって、これを行う。Lombard [4] は、発話者の音声の強さおよび基本周波数（声の高さ・ F_0 ）が、発話中にモニタしている外界ノイズのレベルに合わせて、無意識に大きくなるという現象を観測した。この現象はロンバード効果と呼ばれ、不随意的な反応であることからロンバード反射とも呼ばれている [5]。また、ロンバード効果によって発話者から発された音声はロンバード音声と呼ばれている [6]。例えば、友人と電話で会話をする際、電話越しの声が小さいと自然と自分の声を大きくしてしまう。これは相手側にとって自分の声が大きく聞こえていると錯覚させ、自分の発話音声を小さくするという仕組みである。これは、ロンバード効果によるものである。このように、ロンバード効果は日常的によく観測される現象である。近年では、音声対話システムへの応用に用いられるなど [7, 8, 9]、ロンバード効果については、多くの研究がなされており、雑音の種類によって、この性質が変化することなどが示されてきた [10, 11]。しかし、雑音の種類による話者の発話に対する方略の違いについては明らかにされていない。

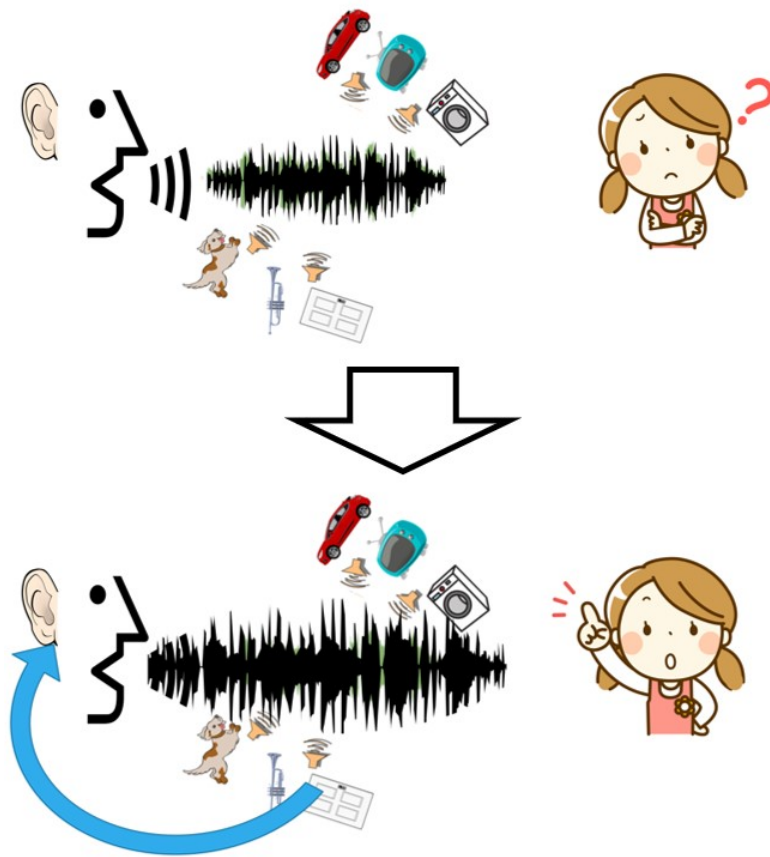


図 1.1: 雑音環境下での発話

1.2 先行研究

外界ノイズのある環境下で音声コミュニケーションを行う際には、聴覚フィードバックの機能から生じるロンバード効果が不随意的に発生する。この現象は、ヒト以外のクジラ・ネコ・カエルといった様々な動物でも観測されている [12]。ロンバード音声は、静かな環境で話された音声と比べて、話速・スペクトル傾斜・強度・フォルマント周波数・F0といった様々な音響特徴の変化が、観測されている [13, 14, 15]。これらの、音響特徴の変化は、背景雑音の音圧レベルに対応している。F0に関しては、周波数の平均・分散ともにロンバード効果とともに値が増加することが知られている [16, 18]。フォルマント周波数に関しては、ロンバード効果によって、第一フォルマント (F1) の周波数が系統的に高くなることが報告されている [11, 19]。また、背景雑音の音圧レベルの増加にともなって連続的に変化することも報告されている。また、第二フォルマント (F2) の周波数に関しては、背景雑音の音圧レベルの増加に伴って、上昇するという報告も存在するが、F1ほど系統的に上昇するわけではないということが報告されている [19]。Stoweら [20] は、帯域制限された広帯域雑音 (0.5–4 kHz) とノッチ雑音を用いて、強度、持続時間、および F0 の変化を調査した。その結果、発話音声と同じ周波数帯域を持つ広帯域雑音の提示下でこれらの値が上昇することを示した。これは、ロンバード効果が雑音の性質によって異なることを示唆している。Matsumotoら [21] は、様々な雑音環境下で発話する際に、発話者が発話を際立たせるためにどのような方略をとっているのかを理解するため、母音の種類の特徴付けに重要な F1・F2 に着目し、雑音の変化に発話者のフォルマント周波数がどのように変化するのか調査を行った。しかし、雑音に応じてフォルマント周波数が増加することが示唆されたものの、傾向をつかむまでには至っていない。

1.3 研究目的

本研究では、母音の認識に重要である F1・F2 [22] に対して、種々の雑音を用いて、音声のフィードバック情報にそれらを付加した際における話者の発話変化の傾向を、聴覚のマスクングパターンを考慮した上で、検討する。これを明らかにすることは、ヒトがどのような方略をとり周囲の雑音環境に合わせて音声を知覚・生成しているのか、そのメカニズムを解明するために重要な検討項目である。

1.4 本論文の構成

本論文は6章で構成される。図 1.2 に本論文の構成を示し、以下に内容を示す。

第1章

雑音環境下における発話に着目した理由について説明し，それについての関連研究について述べ，本研究の目的を明らかにする．

第2章

音声の生成と知覚についてのメカニズムを説明し，それらが一体となる聴覚フィードバックについて述べ，研究の計画を提案する．

第3章

母音発話に影響があり，それによって発話が変化することが期待できる背景雑音の生成手法について説明し，その分析について述べる．

第4章

第3章で作成した雑音を音声のフィードバック情報に付加した際の母音発話がどのように変化したのかについて述べ，また，それらの刺激の聴覚上のスペクトルがどのように表現されたのかについて述べる．

第5章

第4章から，本研究で作成した刺激がどのように母音発話に変化を与えたかについて論じる．また，それが起こった原因や刺激の違いによって母音の発話にどのような変化の傾向が観測されたかについて総合的に論じる．

第6章

本研究で明らかにしたこと，残された課題について述べる．

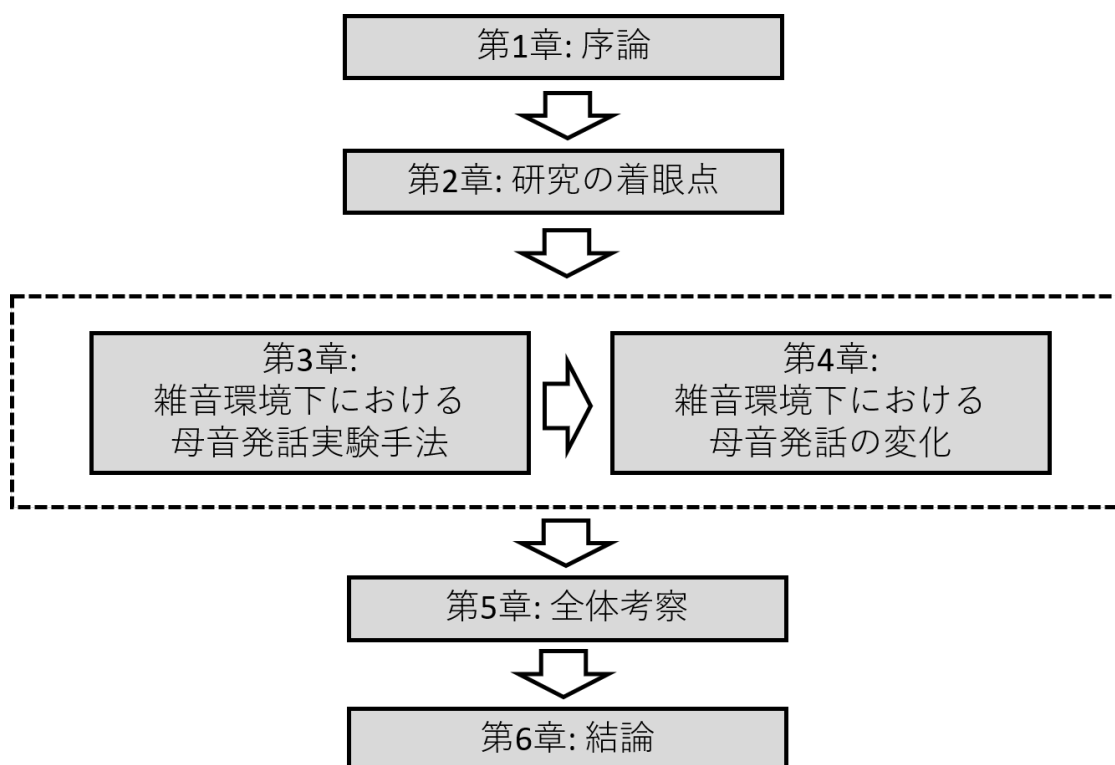


図 1.2: 本論文の構成

第2章 研究の着眼点

2.1 音声生成

音声は、肺、気管、咽頭、咽頭腔、口腔、鼻腔などにより構成される音声器官から空気の振動として生成される [1]. 母音発話の際には、声帯由来の振動音が F_0 となる. この F_0 の違いは、一般に性別で見られ、成人男性の方が成人女性よりも低い. 声帯振動による音源が声道を通ることによって母音は生成される [24]. 声道は、舌、顎、口唇、軟口蓋などの動かすことのできる調音器官と、硬口蓋、咽頭壁などの動かすことのできない声道壁で囲まれ、非常に複雑な三次元形状である [1]. また声道の長さの違いもまた一般に性別で見られ、成人男性の方が成人女性よりも長い. 声道の共鳴特性のことを、フォルマント周波数と呼び、周波数が低い方から F_1 , F_2 と呼ばれる. また、声道の形状を変えることで、音源の特定の周波数（フォルマント周波数）が強められ、/a/, /i/などの母音に対応した音声スペクトルが生成される. 舌の高さと F_1 , 舌の前後方向の位置と F_2 とはそれぞれ相関があり、舌の位置が低くなると F_1 の周波数が高くなり、舌の位置が前になると F_2 の周波数が高くなる [22]. この、 $F_1 \cdot F_2$ は、母音を特徴づける重要なパラメータである. 図 2.1 に、母音毎の $F_1 \cdot F_2$ の分布を示す. このことから、フォルマント周波数の分布で母音を分類できることが分かる. また、母音は、低い周波数の振幅の方が大きい. 以上のことより、低次のフォルマント周波数が発話において重要であることが分かる.

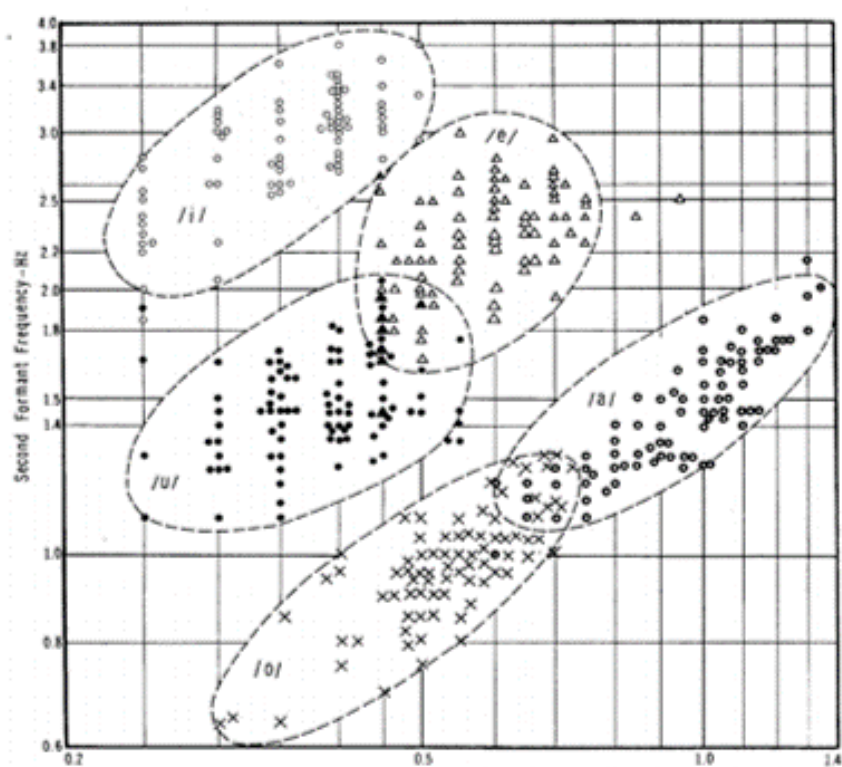


図 2.1: 子供, 若者, 大人が発話した日本語 5 母音の F1・F2 の分布 (参照 [25])

2.2 音声知覚

聴覚器官は大きく外耳，中耳，内耳の3つに分けられる．外耳は，外介と外耳道から成り，鼓膜が中耳との分かれ目である．外耳道は，3-4 kHz 付近に共振特性をもつ．そのため，外耳道入口から鼓膜までの間での音圧比は，低周波部分と比較し 10 dB ほど大きくなっている．中耳は，鼓膜から蝸牛に至るまでの振動伝搬の経路である．中耳は，音波を能率よく内耳に伝えるためのインピーダンスマッチングの役割を担っている．内耳は，蝸牛とよばれ，カタツムリのような螺旋上の形をしている．その内部には，基底膜と呼ばれるものが存在する．これは，音によって振動し，基底膜の振動の振幅が最大となる位置は周波数によって異なる．基底膜の上には，コルチ器に保持された聴神経があり，機械振動が神経発火情報に変換される．この神経発火情報は脳内の様々な場所へ伝達され，周波数分析や音韻処理などといった処理が行われることによって，音声を理解しているといわれている [1]．このことから，脳に伝達されるまでの聴覚系でのスペクトル表現は，物理的なスペクトル表現とは異なるものであることが分かる．これをよく表す例として，聴覚上のマスキングという現象が存在する．ヒトには，聞きたい音とそうでない音を分離できる能力が備わっている [23] が，これが出来ないときにマスキングが起こる．電車の中で友人と会話をしている際に，電車がトンネルに入ると急に会話がしづらくなる．これは，音声によって引き起こされた興奮（神経活動）が，トンネルの中での電車の走行音によって引き起こされた別の興奮によって，マスキングされ聞こえなくなるという聴覚の特性から起こっている．また，周波数マスキングという現象が存在する．この現象は，聴覚の周波数選択性が関係している．基底膜には，周波数を分解する能力があるが，その能力には限界がある．これを反映したのが，聴覚フィルタ [26] という考え方である．聴覚フィルタは，帯域が連続的に重なり合っている帯域通過フィルタであり，それぞれ異なる中心周波数をもつ．このフィルタの特性から図 2.2 に示すように，1 kHz の純音を呈示した場合において，他の周波数帯域にも影響していることが分かる．そのため，物理的なスペクトルで影響しない帯域においても，聴覚的なスペクトル表現では影響があり，マスキングが発生することがある．これを周波数マスキングという．聴覚フィルタの特性上，信号よりも高い周波数への影響が大きい．

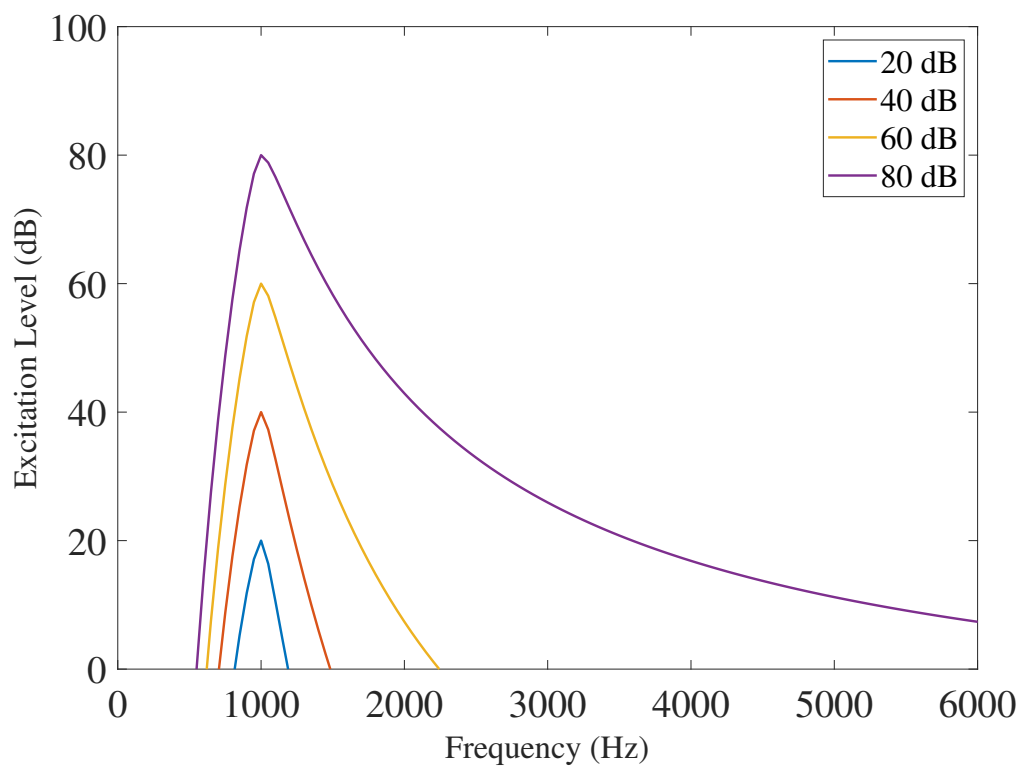


図 2.2: 1 kHz の純音を 20 dB から 80 dB まで 20 dB ずつ付加した際の興奮パターン

2.3 聴覚フィードバック

発話を行う際には音声生成と音声知覚は互いに不可分の関係である。聴覚フィードバックとは、話者が自分の声を自分自身で聞き取りながら、発話を制御するためのフィードバックのことで、これによって発話を正確に行うことが出来ると言われている。発話運動制御メカニズムの概念を図 2.3 に示す。聴覚フィードバック研究で使われるテクニックは、発話中のフィードバック情報に何らかの外乱（摂動）を加え、発話運動がどのように変化するかを観測し、そこから発話運動制御系のある側面を同定しようとするものである [27]。本研究でも、このテクニックを使用し、母音発話時に話者の音声フィードバック情報に、母音の発話に重要であると言われている $F1 \cdot F2$ に関連した刺激を付加した際の発話変動を観測することで、発話運動制御系の一面を同定することを試みる。

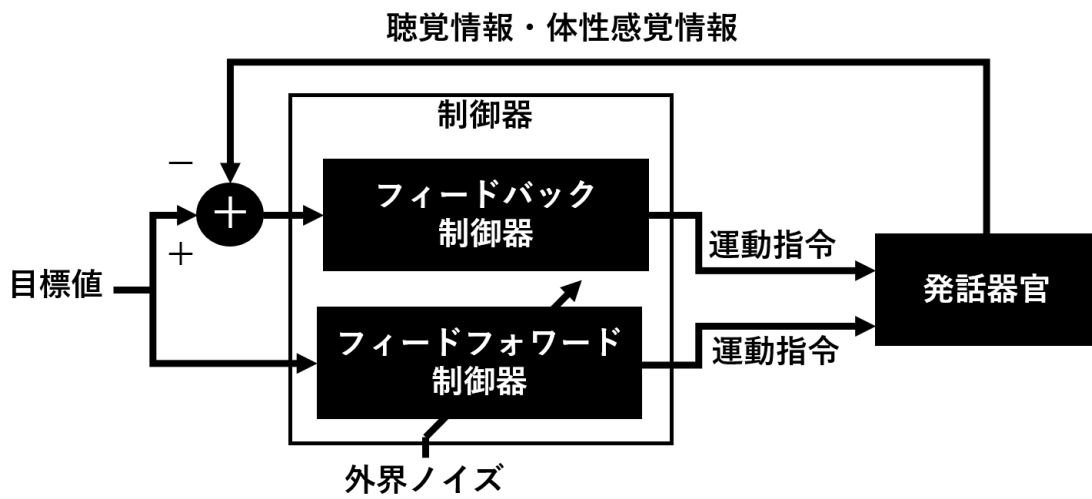


図 2.3: 発話運動メカニズムの概念 (参照 [1])

2.4 研究計画

本研究では、背景雑音の種類に依存とされている発話の変化の傾向をつかむために、母音の知覚に重要な F1・F2 に影響のある刺激のパターンを 7 種類用意し、それぞれ音圧レベルを変化させて母音発話中の話者の音声フィードバック情報に付加し、その音声を観測する。その後、その音声を静かな環境で発話された母音の音声と比較することで、発話の変化の傾向をつかむことを試みる。そのための、研究の手続きを説明する。まず、実験刺激を作成するために、静かな環境での話者の音声 (/a/, /i/, /u/, /e/, /o/) を収録し、それぞれの F1・F2・F3 推定値を算出する。その後、算出されたフォルマント周波数を元に各刺激を話者・母音毎に作成する。作成された刺激を母音発話中の話者に呈示し、その音声を収録する。そこから、F1・F2 の周波数および振幅の変化を、静かな環境で発話された音声のものと比較する。最後に、それらの刺激が付加された際の聴覚上のスペクトル形状を考慮することで雑音の種類による発話の変化の傾向をつかむことを試みる。

第3章 雑音環境下における母音発話 実験手法

3.1 概要

本研究は、雑音の種類による話者の発話の方略を調査し、その傾向をつかむため、発話に重要な F1・F2 に焦点をあて、そこに対する刺激を数種類提示し、発話された音声を取録する。発話内容は、日本語 5 母音 /a/, /i/, /u/, /e/, /o/ である。調査するパラメータは、フォルマント周波数および振幅、刺激の神経興奮パターンである。これを調査することで、雑音環境下における母音発話の口の構えと、聴覚上でのスペクトル形状を考慮することができる。そのため、生成と知覚の両方の視点から話者の方略を検討する。

3.2 実験刺激

実験に使用する刺激は、低域通過雑音 (LPN)、高域通過雑音 (HPN)、帯域通過雑音 (BPN)、帯域阻止雑音 (NN)、およびピンク雑音 (PN) である。これらの刺激は式 (3.1) に基づいて生成する。このとき、 N は生成される刺激、 f_L は生成される刺激の周波数帯域の下限、 f_U は生成される刺激の周波数帯域の上限、 f は任意の周波数、 T は時間長、 t はサンプル時間系列、そして $\phi(f)$ はランダムイズ位相を示している。この $\phi(f)$ の f は各刺激の帯域に対応して変化する。帯域阻止雑音は、低域通過雑音と高域通過雑音を合わせることで表現される。提示する刺激は、フォルマント周波数の分析結果に基づいて、周波数帯域を各発話者のフォルマント周波数と母音の種類に対応するように生成された。また振幅は、まず下記の式 (3.1) に基づいて生成された広帯域雑音 (1-22,050 Hz) を、RMS 値で 70 dB, 80 dB となるように校正し、その後、広帯域雑音と dB/Hz が同じになるように調整した。また、本研究では、実験環境の有効性およびロンバード効果が発生するかどうかを検証するため、先行研究と同様の雑音である PN を用いて発話の変化を調査した。実験で用いた刺激を、図 3.1-4 に示す。PN は、広帯域白色雑音に高域減衰型のフィルタを適用して作成した。そのためこの雑音は、低域が強調された雑音であることが図 3.1 から分かる。PN の音圧レベルは、RMS 値で 70 dB, 80 dB となるように調整された。

$$N = \sum_{f=f_L}^{f_U} \sin\left(2\pi f \frac{t}{T} + \phi(f)\right) \quad (3.1)$$

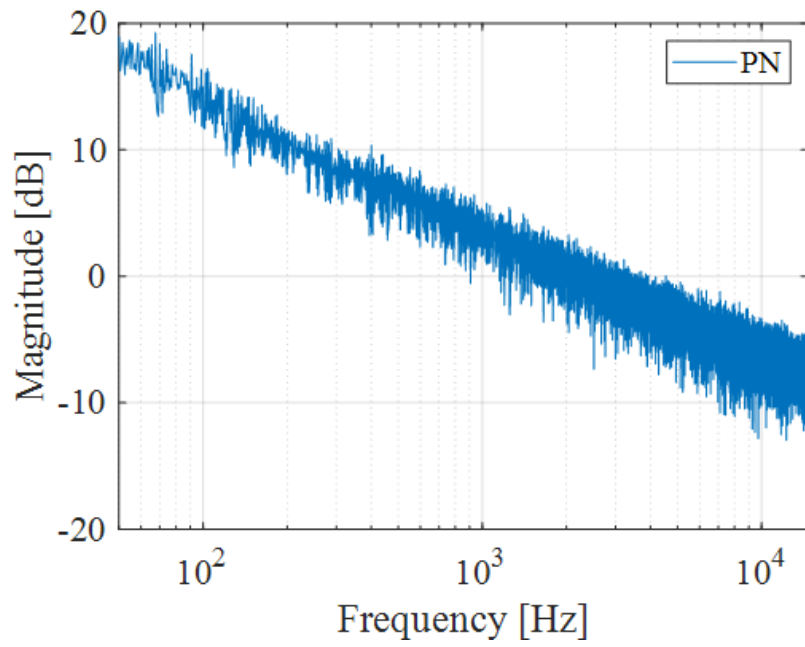


図 3.1: 実験刺激 (PN)

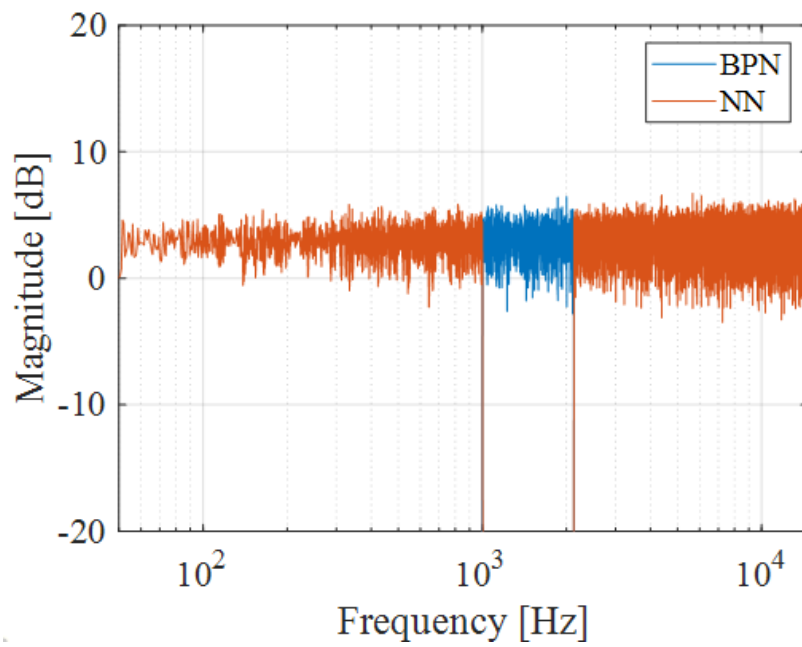


図 3.2: 実験刺激 (BPN, NN)

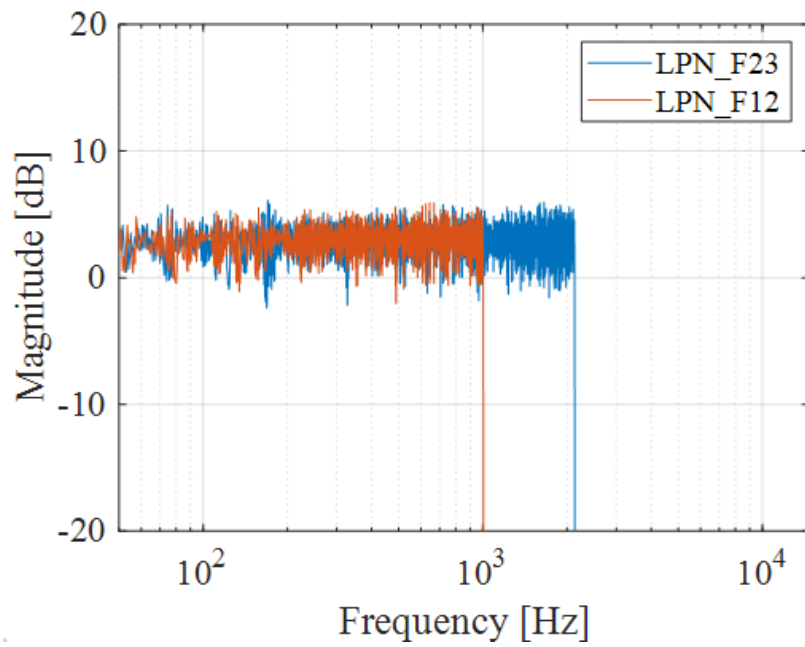


図 3.3: 実験刺激 (LPN_F12, LPN_F23)

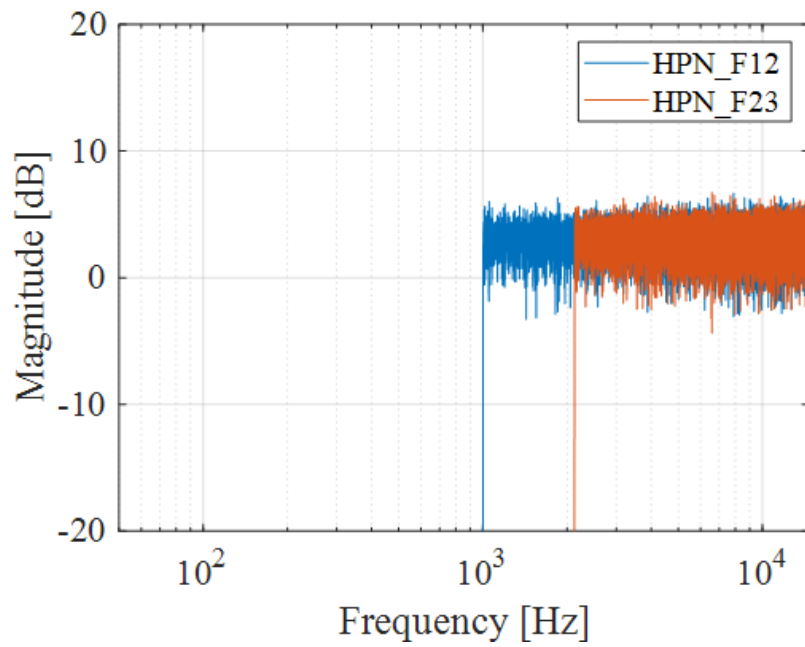


図 3.4: 実験刺激 (HPN_F12, HPN_F23)

3.3 静かな環境での発話の収録と分析

静かな環境における発話と雑音環境下における発話の変化を調査するため、被験者4名が実験に参加した。実験参加者は、北陸先端科学技術大学院大学の22-23歳の学生4名（男性2名、女性2名）であった。被験者は、母国語が日本語であり、日常生活に支障のない程度の聴力を有していた。また話者の母音毎に対応した刺激を生成するため、発話者には、タスクとして日本語の母音を、/a/, /i/, /u/, /e/, /o/の順で各5試行ずつ発話するように指示した。この時、母音1種類あたり5個の音声データを得た。また、サンプリング周波数は44,100 Hzとした。この時の実験環境を3.5に示す。ヘッドフォン（STAX SR-L500）はオープンエアタイプであり、この実験ではヘッドフォンから刺激は提示しない。収録した母音のフォルマント周波数を、LPCに基づいた音声分析ソフトであるAcoustic coreを使用し、算出した[28]。このとき、母音の発話区間の定常部を分析区間として使用した。そこから、各発話者のF1, F2, F3の推定値を算出した。表3.1-4に各発話者のフォルマント周波数の分析結果を示す。表中の数値は、母音毎の分析結果の平均値である。また、聴覚における実験刺激のフォルマントへの影響を考慮するため、先行研究[30, 31]に基づいて、神経興奮パターン[29]を算出した。

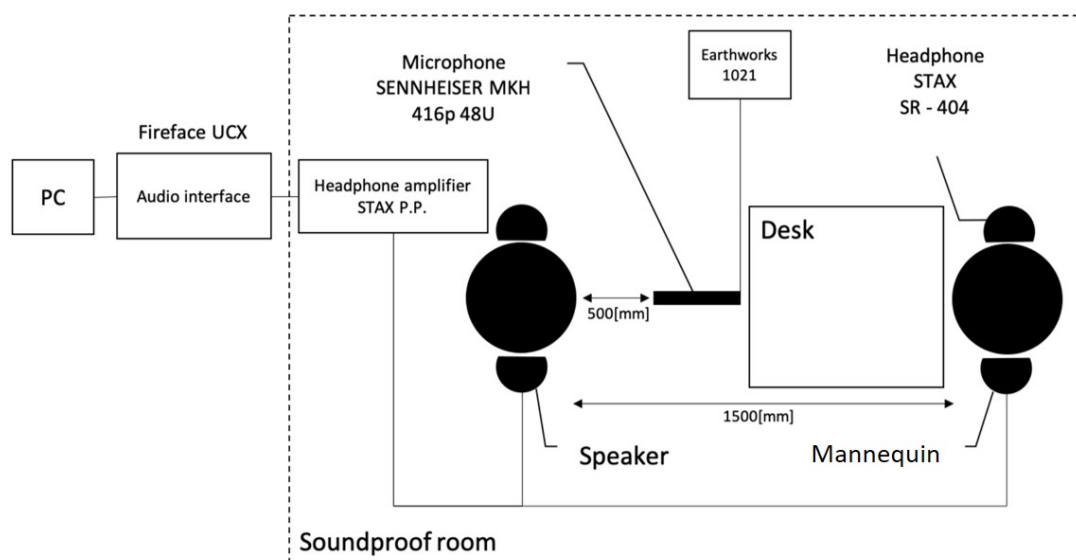


図 3.5: 実験環境

表 3.1: 被験者 1 における平静発話時の母音のフォルマント周波数 (成人男性)

母音	F1 [Hz]	F2 [Hz]	F3 [Hz]
/a/	778	1,224	3,029
/i/	332	2,569	3,366
/u/	371	1,471	2,602
/e/	460	2,128	2,874
/o/	457	827	3,089

表 3.2: 被験者 2 における平静発話時の母音のフォルマント周波数 (成人男性)

母音	F1 [Hz]	F2 [Hz]	F3 [Hz]
/a/	630	1,075	2,883
/i/	329	2,226	3,064
/u/	348	1,275	2,380
/e/	459	1,908	2,655
/o/	437	678	2,863

表 3.3: 被験者 3 における平静発話時の母音のフォルマント周波数 (成人女性)

母音	F1 [Hz]	F2 [Hz]	F3 [Hz]
/a/	715	1,231	3,143
/i/	379	2,609	3,395
/u/	379	1,637	2,559
/e/	455	2,108	2,998
/o/	479	745	2,958

表 3.4: 被験者 4 における平静発話時の母音のフォルマント周波数 (成人女性)

母音	F1 [Hz]	F2 [Hz]	F3 [Hz]
/a/	877	1,399	3,144
/i/	462	2,579	3,461
/u/	459	1,854	2,781
/e/	582	2,195	3,110
/o/	537	983	2,926

第4章 雑音環境下における母音発話の変化

4.1 フォルマント周波数に対応した刺激の生成

フォルマント周波数の分析結果から，呈示する刺激の周波数帯を，話者または母音ごとに決定した． f_L は F1 と F2 の中間周波数に， f_U は F2 と F3 の中間周波数に決定された．これらの値は，刺激の種類に依存する．

4.2 雑音環境下における発話の収録

雑音環境下におけるフォルマント振幅および周波数の変化を調査するため，図 3.5 のヘッドフォンから，刺激が呈示された．呈示された刺激は，話者それぞれの母音に対して，BPN（カットオフ周波数: F1-F2 の中間周波数から F2-F3 の中間周波数），LPN_F12（カットオフ周波数: F1-F2 の中間周波数），LPN_F23（カットオフ周波数: F2-F3 の中間周波数），HPN_F12（カットオフ周波数: F1-F2 の中間周波数），HPN_F23（カットオフ周波数: F2-F3 の中間周波数），BPN（カットオフ周波数: F1-F2 の中間周波数から F2-F3 の中間周波数），PN の順で提示した．これらの実験刺激は，70 dB，80 dB の順で呈示された．またタスクは，3.3 と同様に，各刺激ごとに 5 つの音声データを取得するものであり，各刺激は 20 秒間呈示された．そのため，このタスクは母音ごとに 14 タスク（7 つの刺激 x 2 つの音圧レベル）用意され，合わせて 70 タスクが行われた．この実験では被験者ごとに，350 個の音声データが得られた．そのため本実験では，1200 個の音声を分析した．図 4.1 に，被験者 1 の /a/ に 70 dB の BPN を音声のフィードバック情報に付加した際の物理的なスペクトルを示す．また，図 4.2 にその時の聴覚上のスペクトル形状を示す．この結果からも分かるように，2.2 で示した聴覚特性が考慮されていることが分かる．

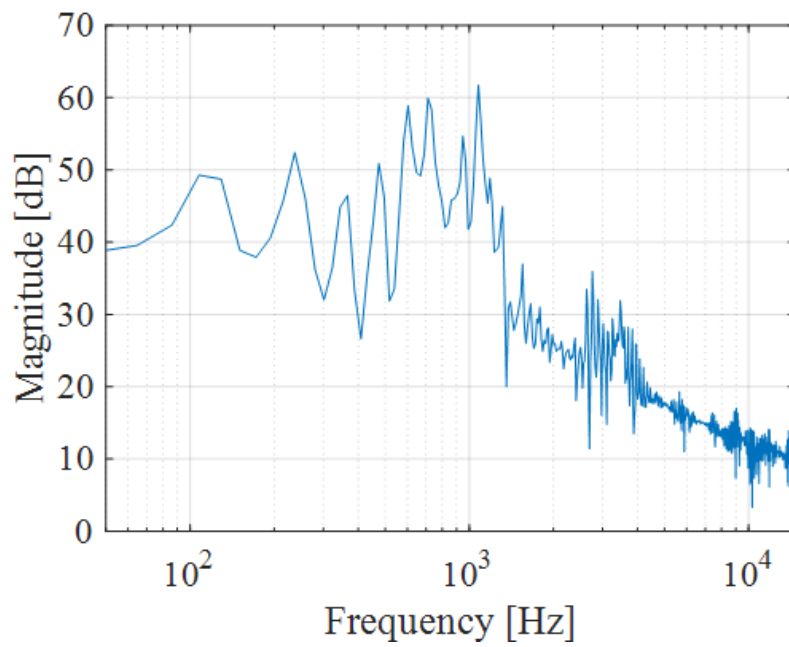


図 4.1: 音声フィードバック情報に 70 dB の BPN を付加した際のスペクトル (被験者 1, /a/)

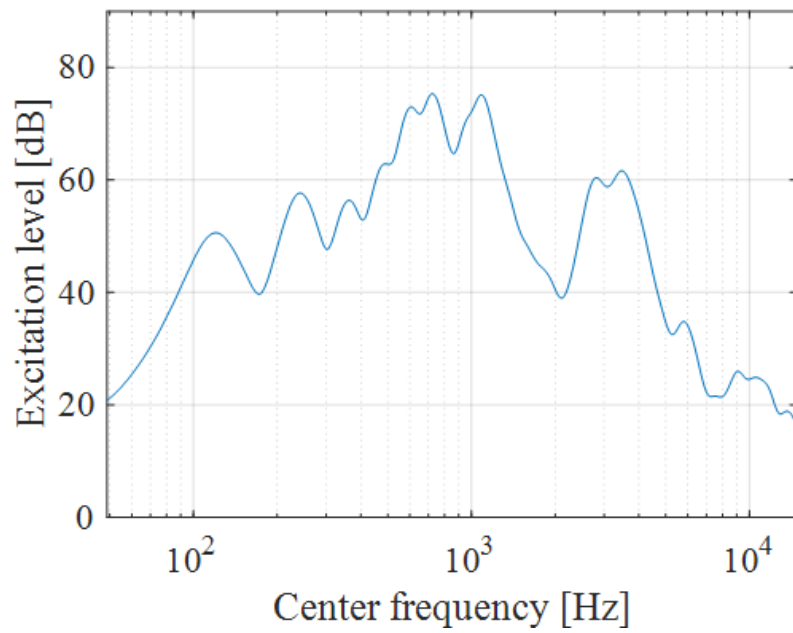


図 4.2: 音声フィードバック情報に 70 dB の BPN を付加した際の興奮パターン（被験者 1, /a/）

4.3 雑音環境下におけるフォルマントの変化

4.3.1 結果

雑音環境下における母音発話と、静かな環境下における母音発話を比較、分析するため、フォルマント周波数および振幅を調査した。図 4.3-16 に実験刺激毎の F1・F2 の変化を示す。フォルマント周波数は、静かな環境における母音発話の F1・F2 の周波数をそれぞれ基準とし、雑音環境下における母音発話がそこから何%変化したかを表している。横軸に F1 の変化を、縦軸に F2 の変化を表している。フォルマント振幅は、静かな環境における母音発話の F1・F2 の振幅をそれぞれ基準とし、そこから何 dB 変化したかを表している。横軸に F1 の変化を、縦軸に F2 の変化を表している。これらの結果から、被験者によってまた母音ごとによって、音圧レベルの差による変化の傾向が似ているということが分かる。また、実験刺激の種類によって、変化の大きさに違いがあることが分かる。この結果から、実験刺激による違いを検討するため、これらの変化を母音・話者で平均し、実験刺激と音圧ごとに表したものが図 4.17 である。静かな環境での発話を基準とし、そこから F1・F2 の周波数が何%変化したか、また、振幅が何 dB 変化したかをそれぞれ表している。横軸は、実験刺激の音圧レベルである。この結果から、F1 は主に周波数方向に、F2 は主に振幅方向に変化する傾向があることが分かった。また、背景雑音として PN を付加した場合、発見された傾向の変化が、他の雑音を付加した場合と比べて、最も大きいことが分かった。一方、背景雑音として HP_F23 を付加した場合、発見された傾向の変化は、他の雑音を付加した場合と比べて最も小さいことが分かった。また、背景雑音として、PN を付加した場合と、NN を付加した場合において、F1 の周波数の変化は、どちらの実験刺激の場合においても同じような傾向を示していることが分かる。しかし、フォルマントの振幅の変化に関しては、80 dB を付加した際において、PN の場合の方が NN と比べて大きいことが分かる。LPN の場合と HPN の場合を比較すると、発見した傾向の変化は、LPN を付加した際において、HPN を付加した際に比べて大きいことが分かる。

4.3.2 考察

雑音環境下において、ヒトの発話は、雑音の種類に関わらず、F1 は周波数方向に、F2 は主に振幅方向へ変化するという傾向が得られた。そのため、音声生成の過程で、F1・F2 を操る舌の運動に関しては、雑音の種類に関わらず、その音圧レベルによって、下方に遷移するということが推察される。前後方向の運動に関しては、雑音の種類またはその音圧レベルに関わらず、舌の位置が一定の方向へ遷移するという知見は得られなかった。逆に言うと、雑音の種類や、音圧レベルに関わらず、発話器官が音声生成する際の運動指令には、舌の前後方向への運動に関して定まった指令が無いということが明らかになったと言える。また、音声の

強度に関しては、雑音環境下においては、F2の振幅を主に大きくするように調音器官が働くということが推察される。

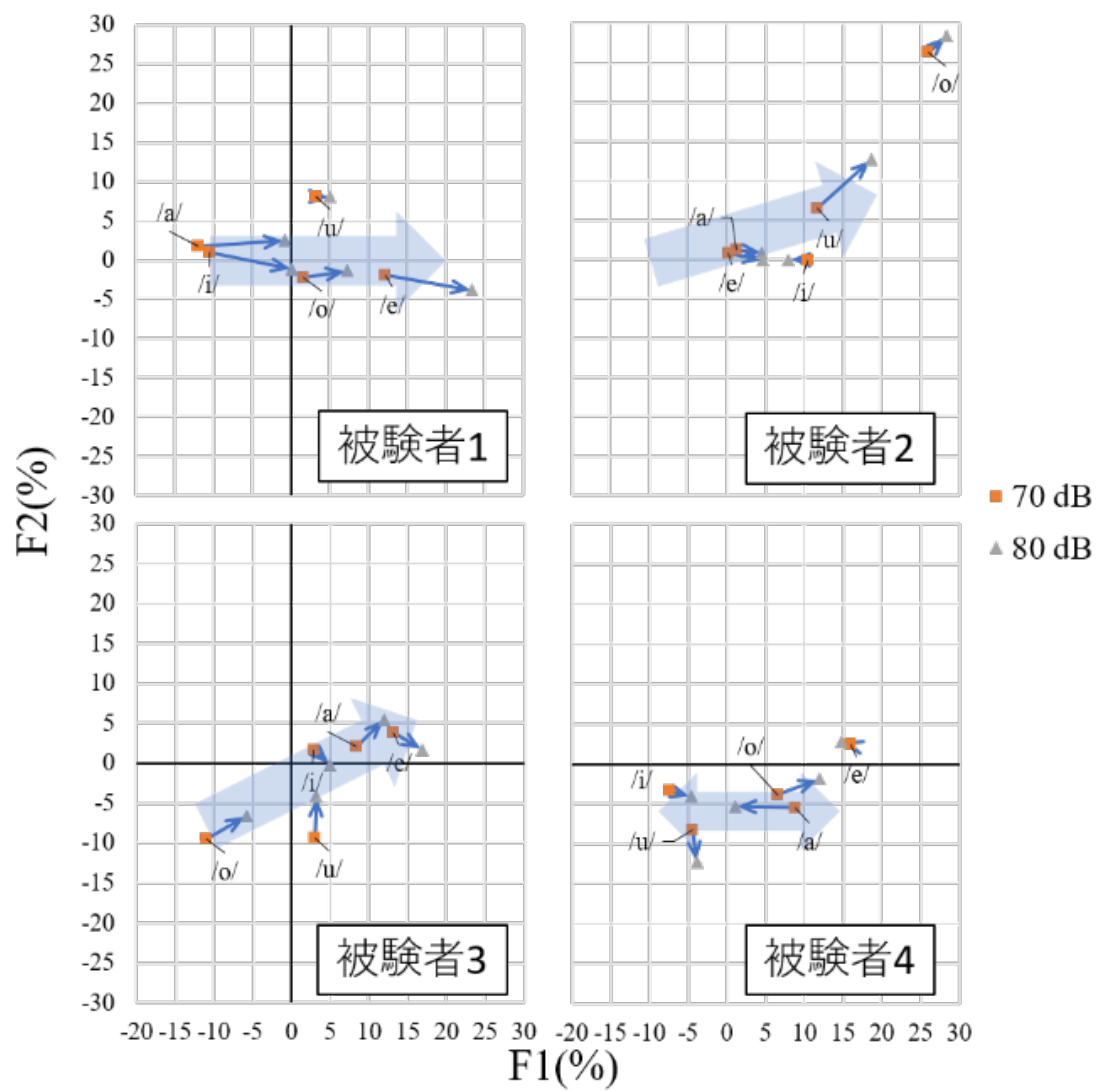


図 4.3: 背景雑音としてピンク雑音を付加した際の被験者毎のフォルマント周波数の変化

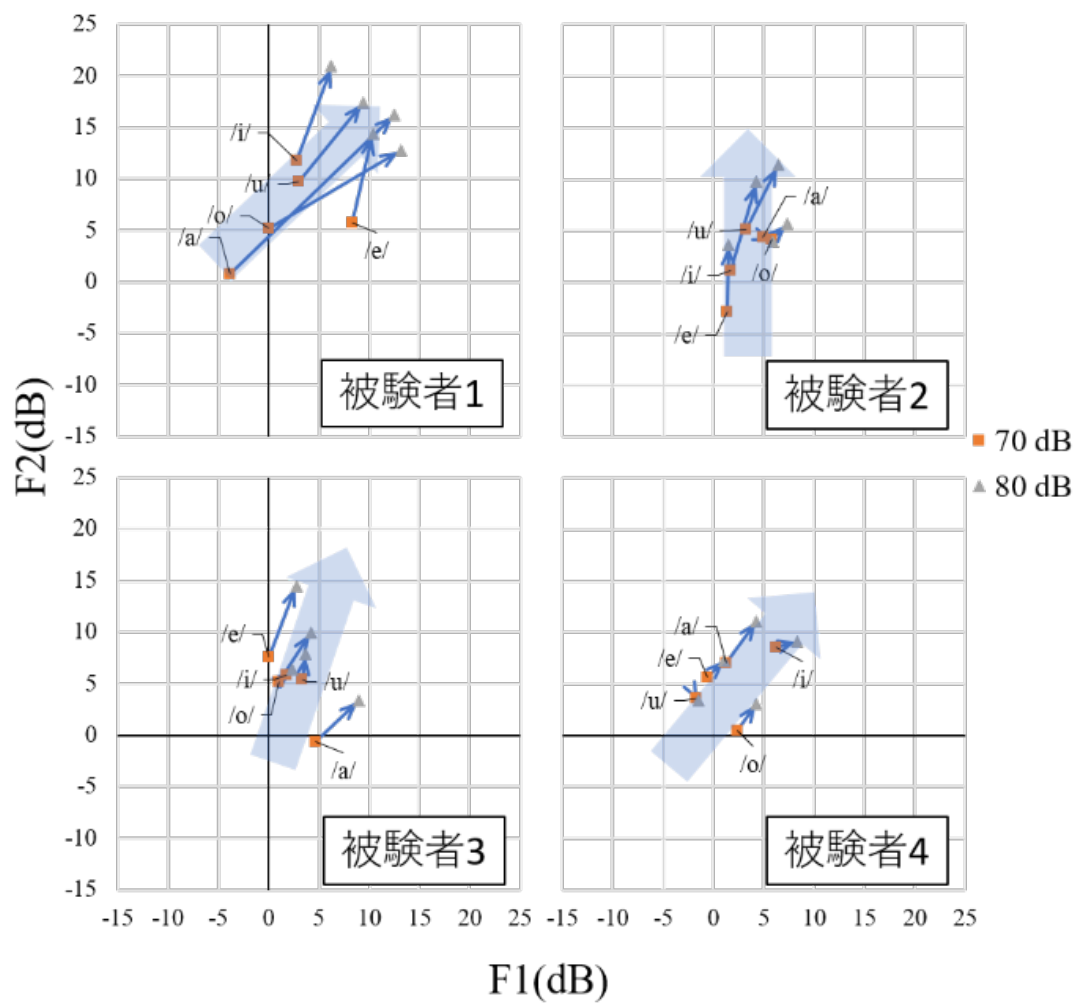


図 4.4: 背景雑音としてピンク雑音を付加した際の被験者毎のフォルマント振幅の変化

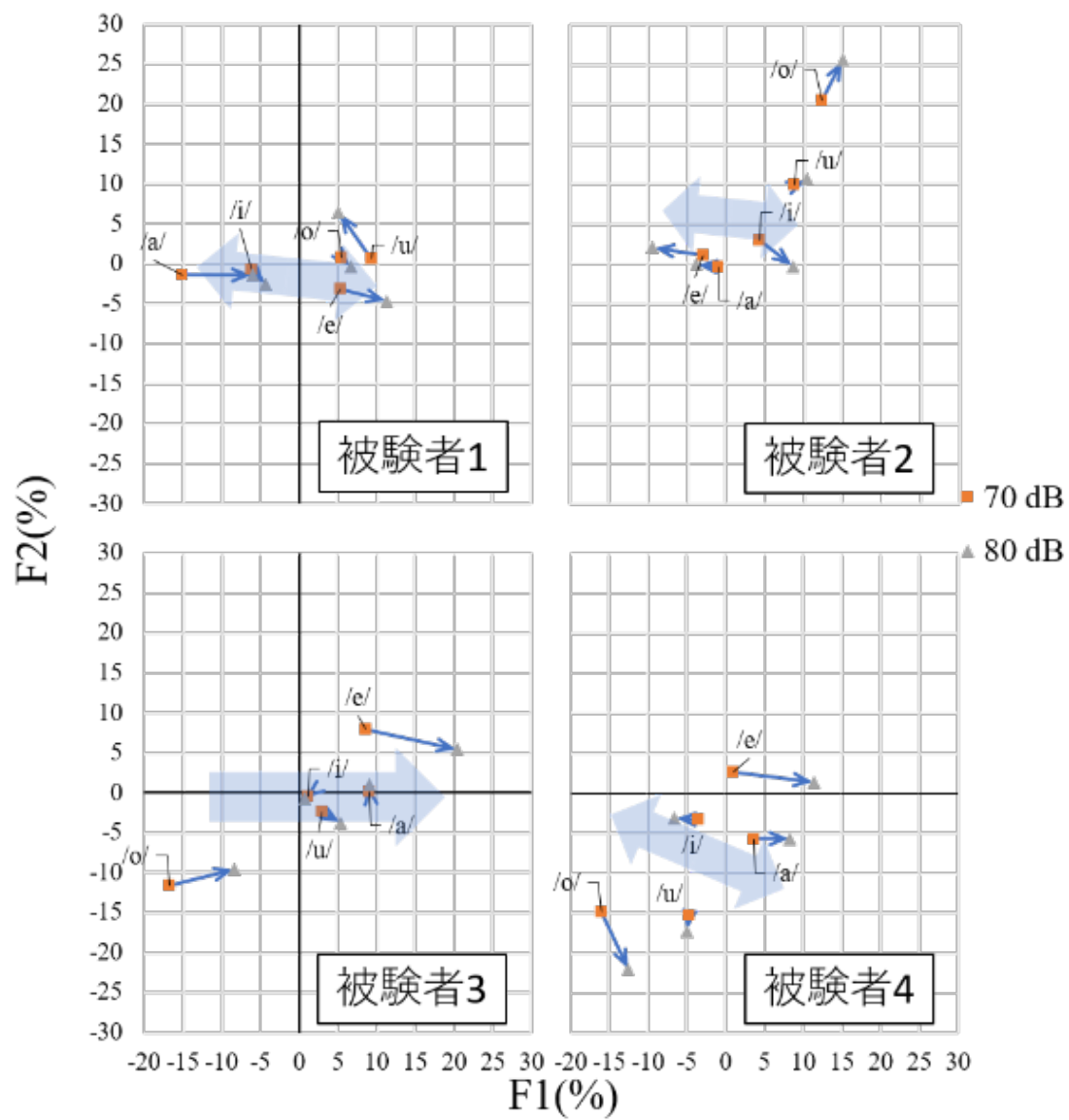


図 4.5: 背景雑音として LPN.F12 を付加した際の被験者毎のフォルマント周波数の変化

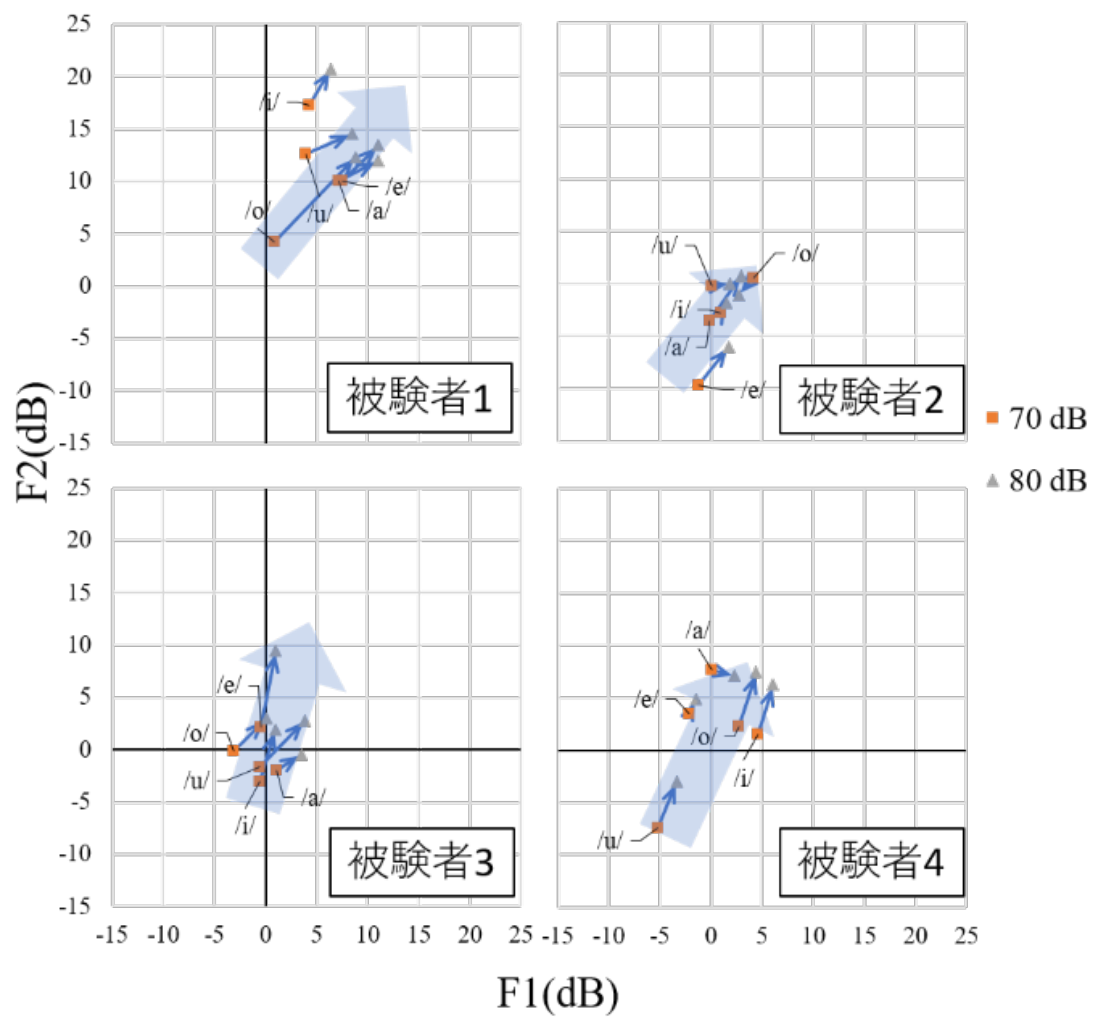


図 4.6: 背景雑音として LPN.F12 を付加した際の被験者毎のフォルマント振幅の変化

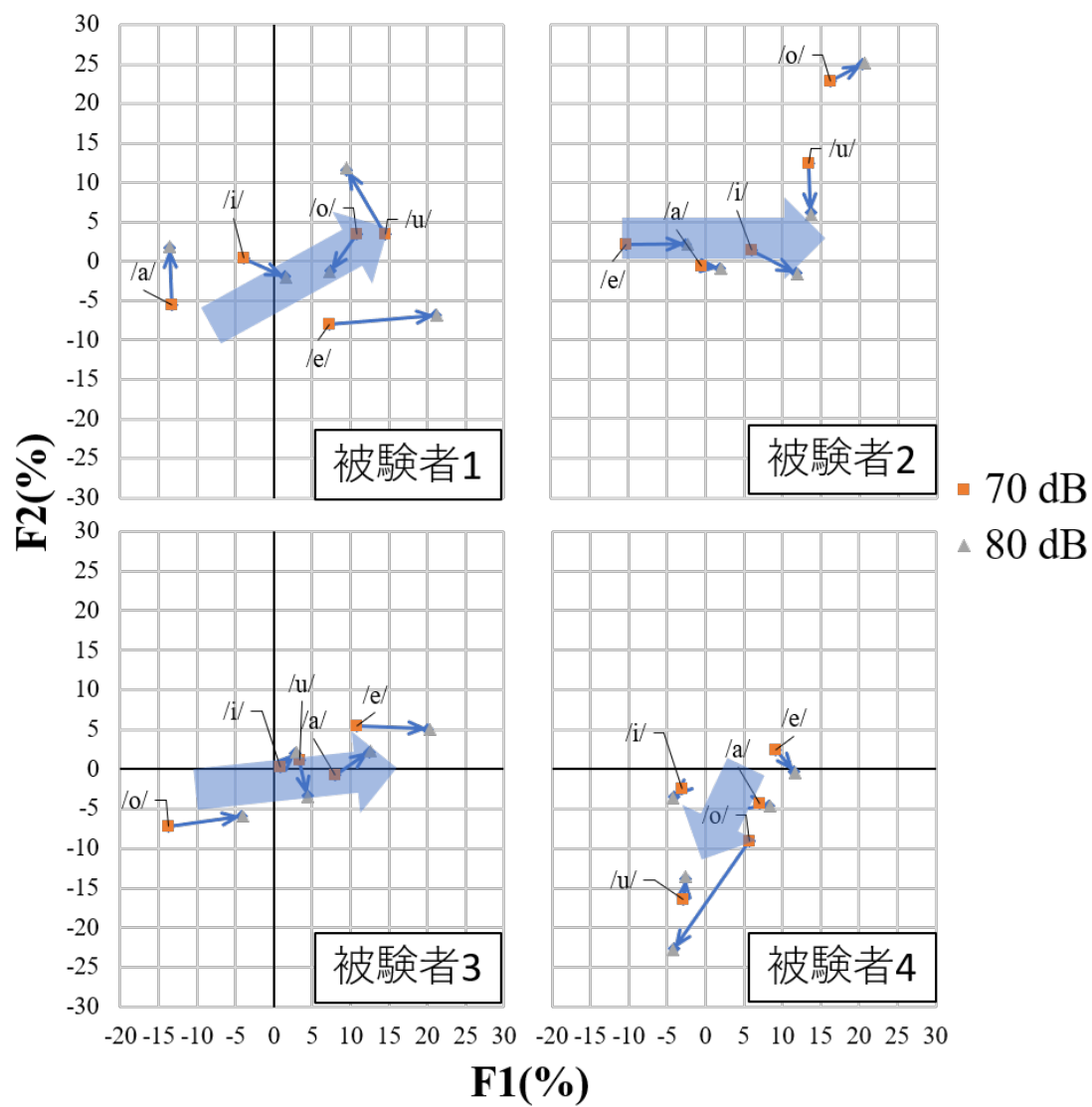


図 4.7: 背景雑音として LPN_F23 を付加した際の被験者毎のフォルマント周波数の変化

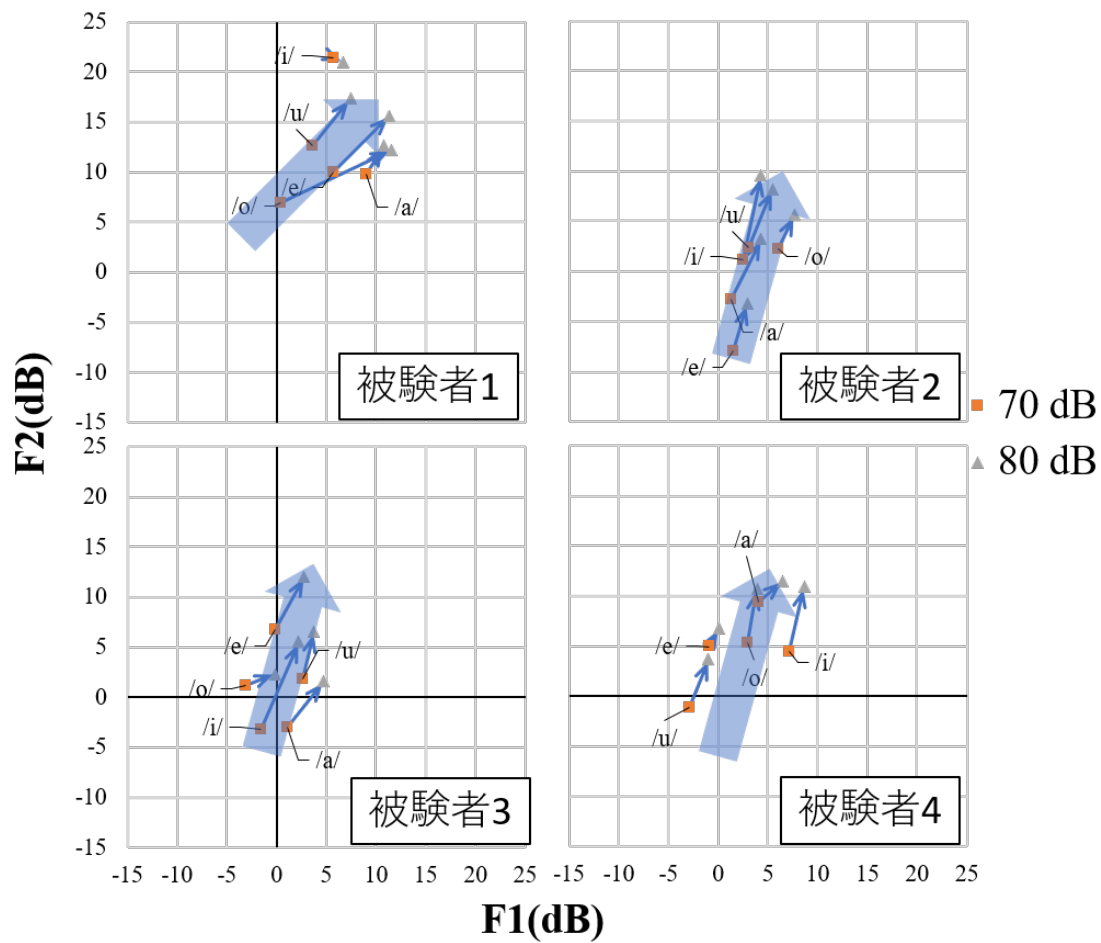


図 4.8: 背景雑音として LPN_F23 を付加した際の被験者毎のフォルマント振幅の変化

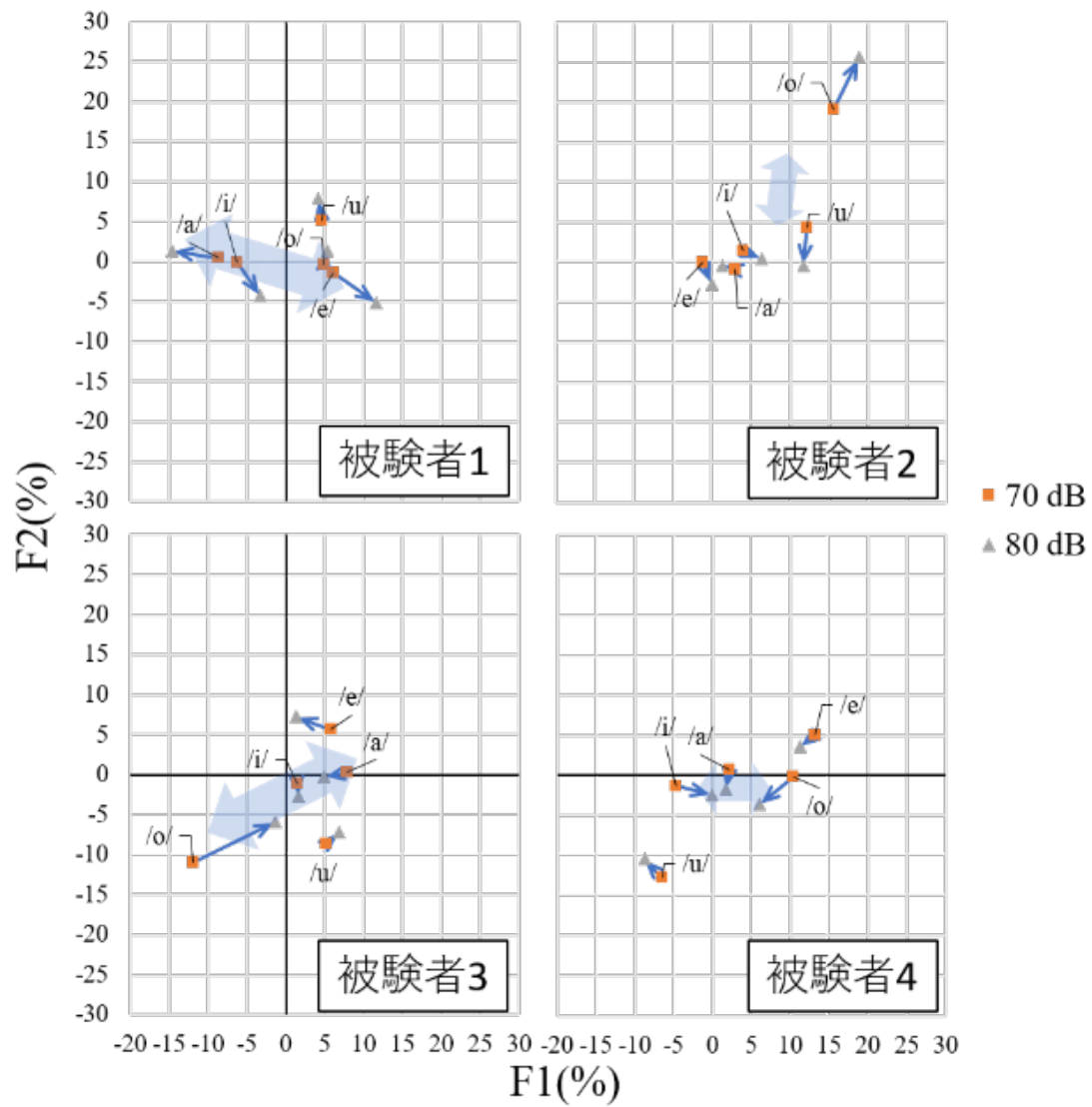


図 4.9: 背景雑音として HPN_F12 を付加した際の被験者毎のフォルマント周波数の変化

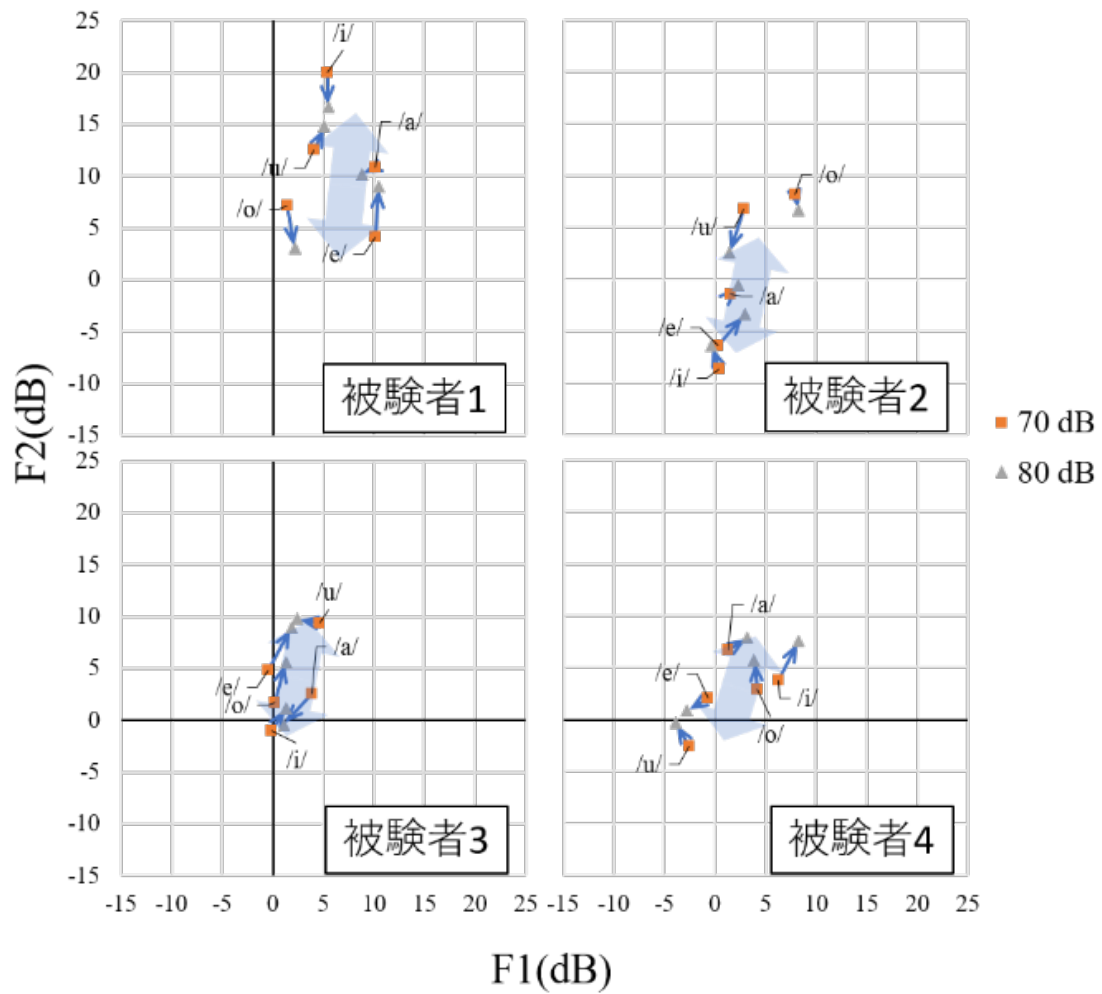


図 4.10: 背景雑音として HPN_F12 を付加した際の被験者毎のフォルマント振幅の変化

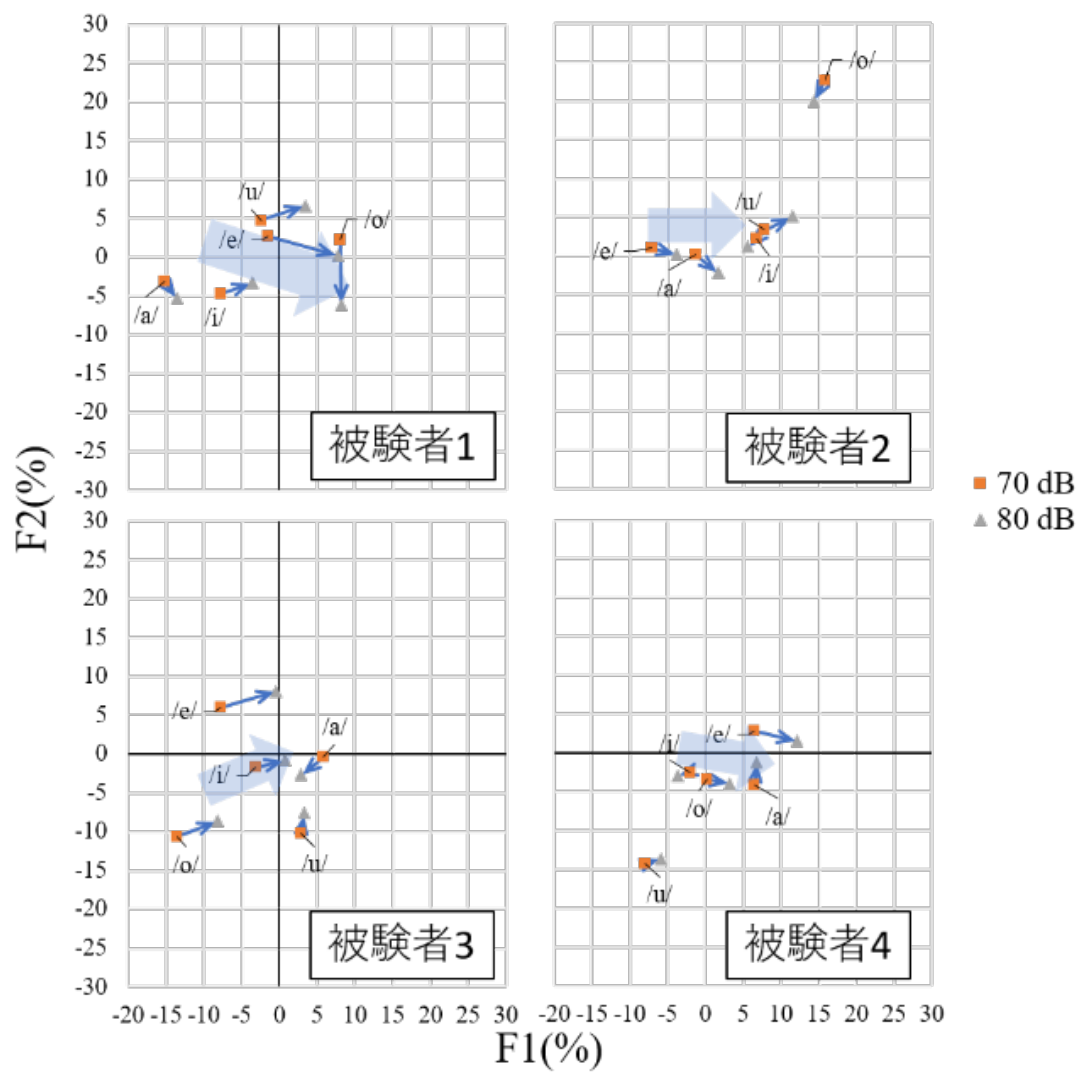


図 4.11: 背景雑音として HPN_F23 を付加した際の被験者毎のフォルマント周波数の変化

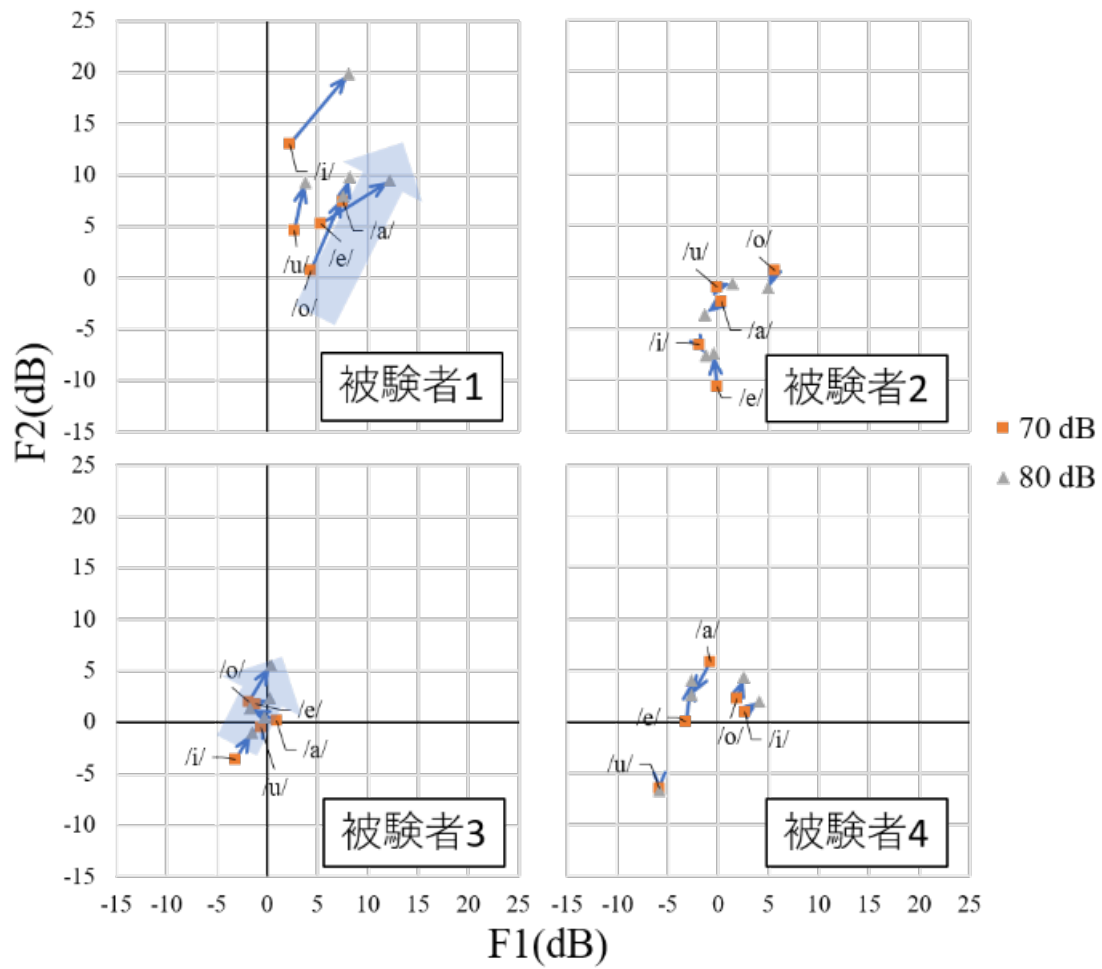


図 4.12: 背景雑音として HPN_F23 を付加した際の被験者毎のフォルマント振幅の変化

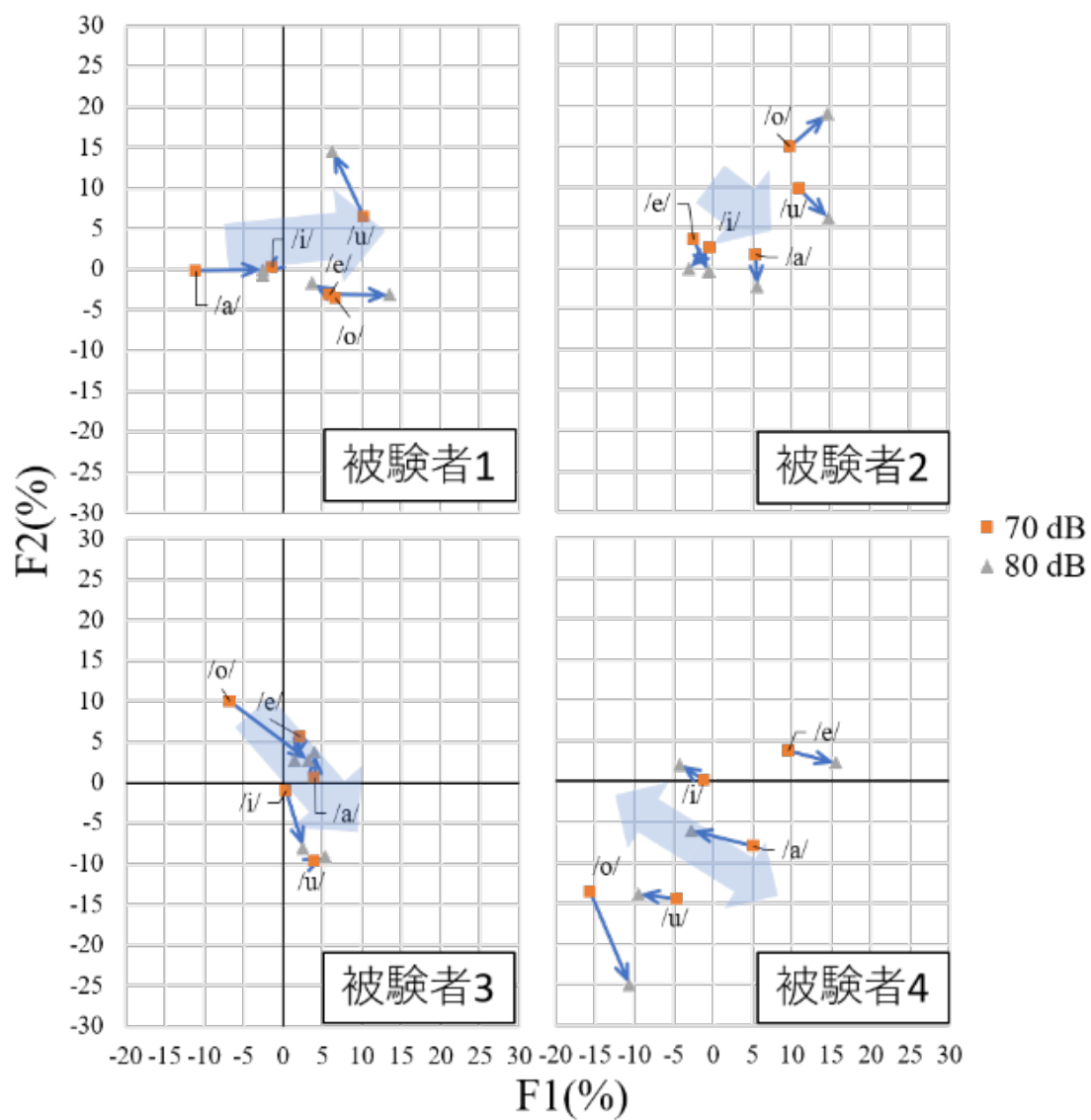


図 4.13: 背景雑音として NN を付加した際の被験者毎のフォルマント周波数の変化

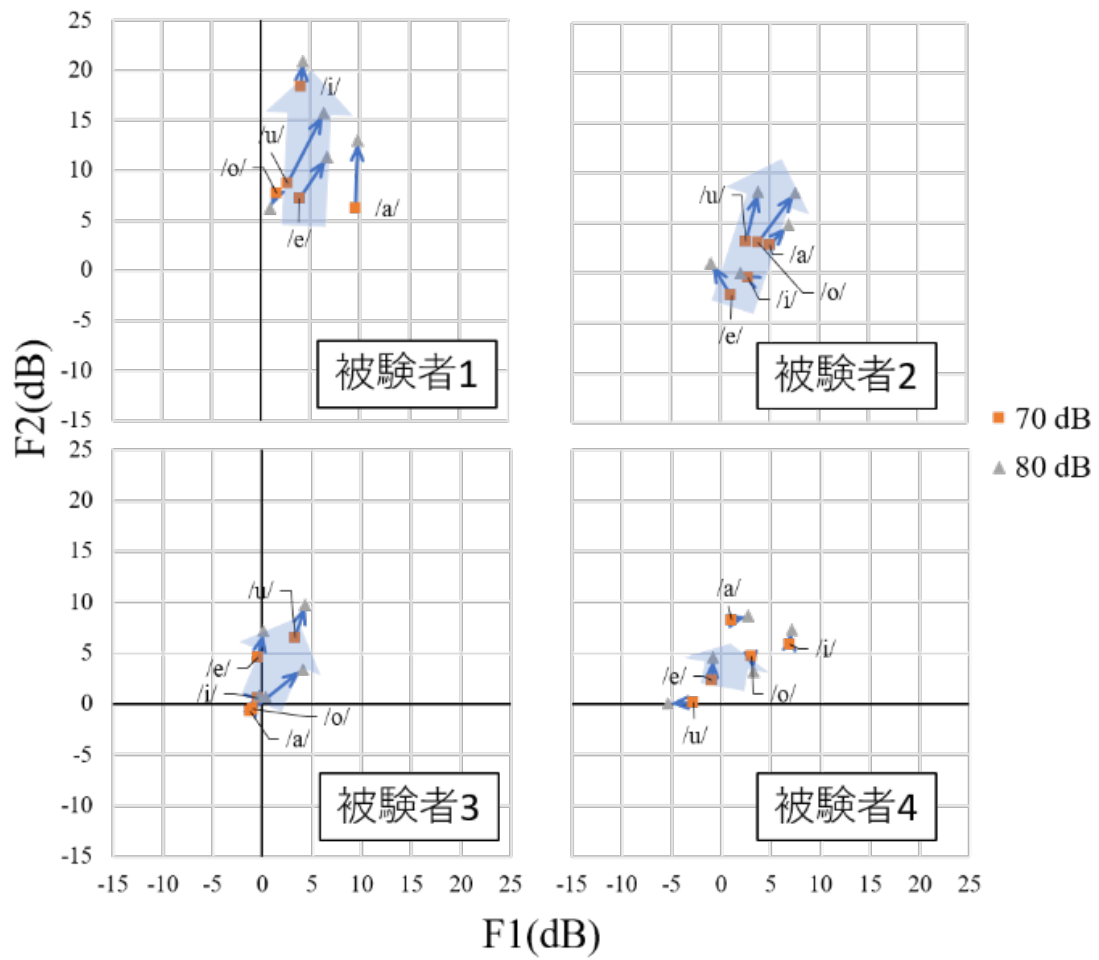


図 4.14: 背景雑音として NN を付加した際の被験者毎のフォルマント振幅の変化

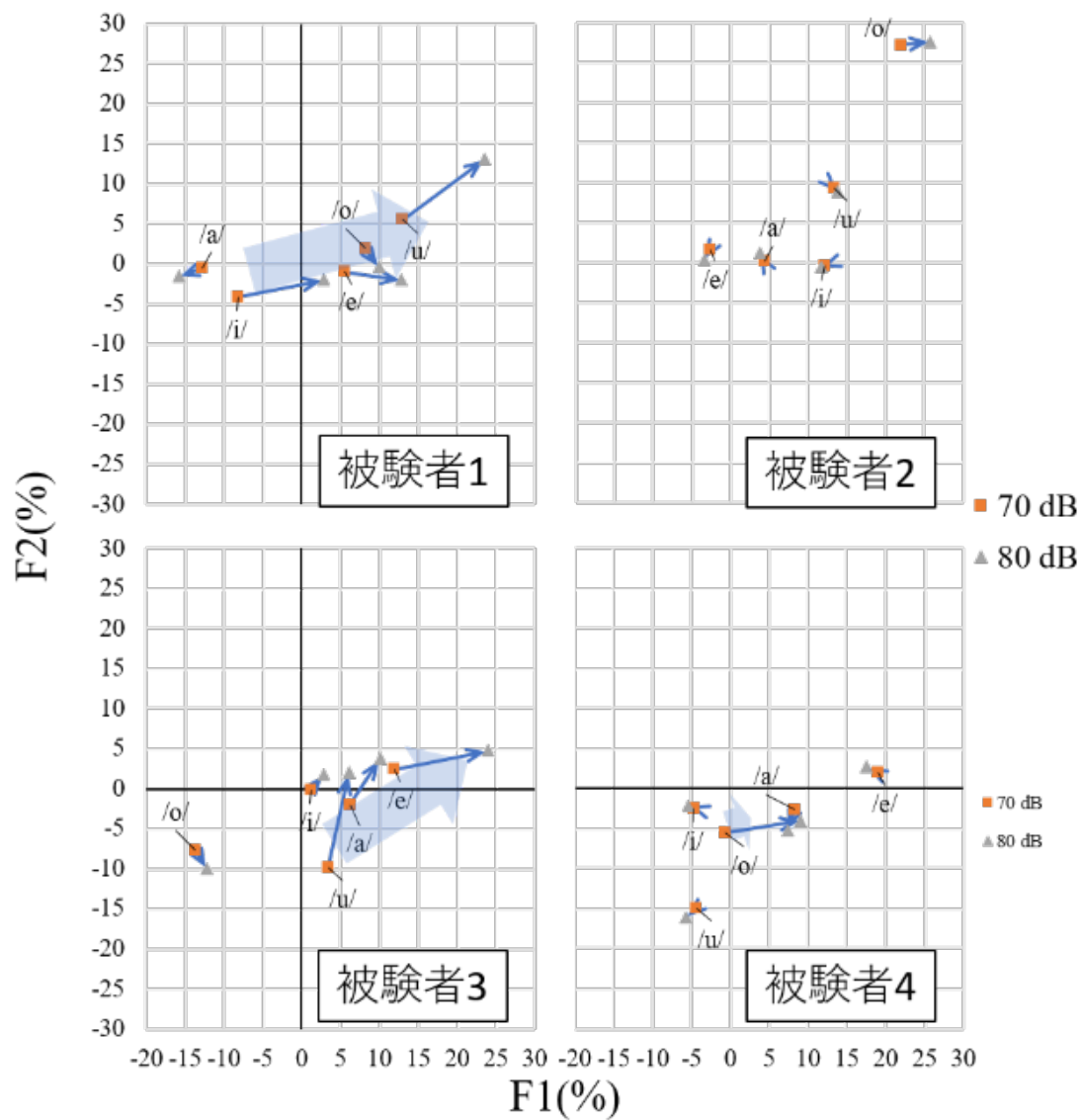


図 4.15: 背景雑音として BPN を付加した際の被験者毎のフォルマント周波数の変化

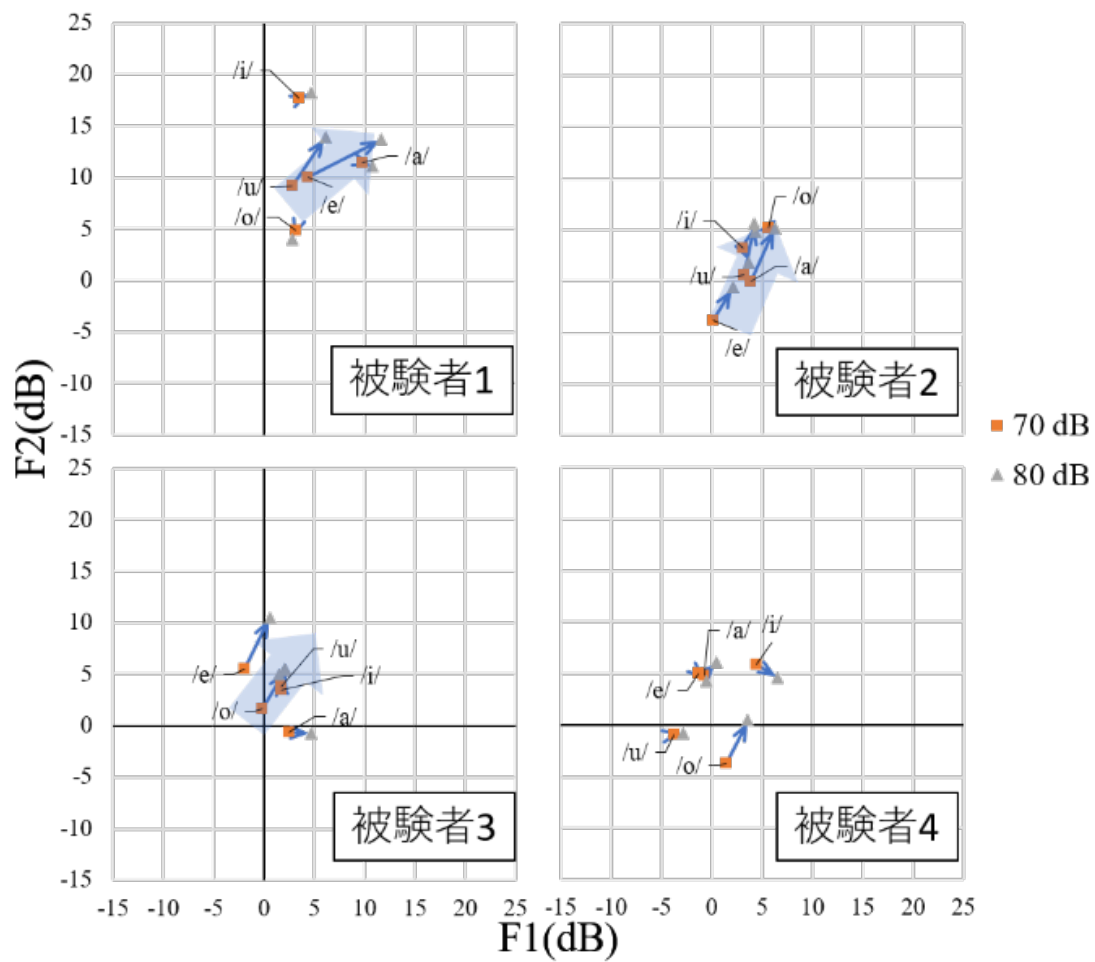


図 4.16: 背景雑音として BPN を付加した際の被験者毎のフォルマント振幅の変化

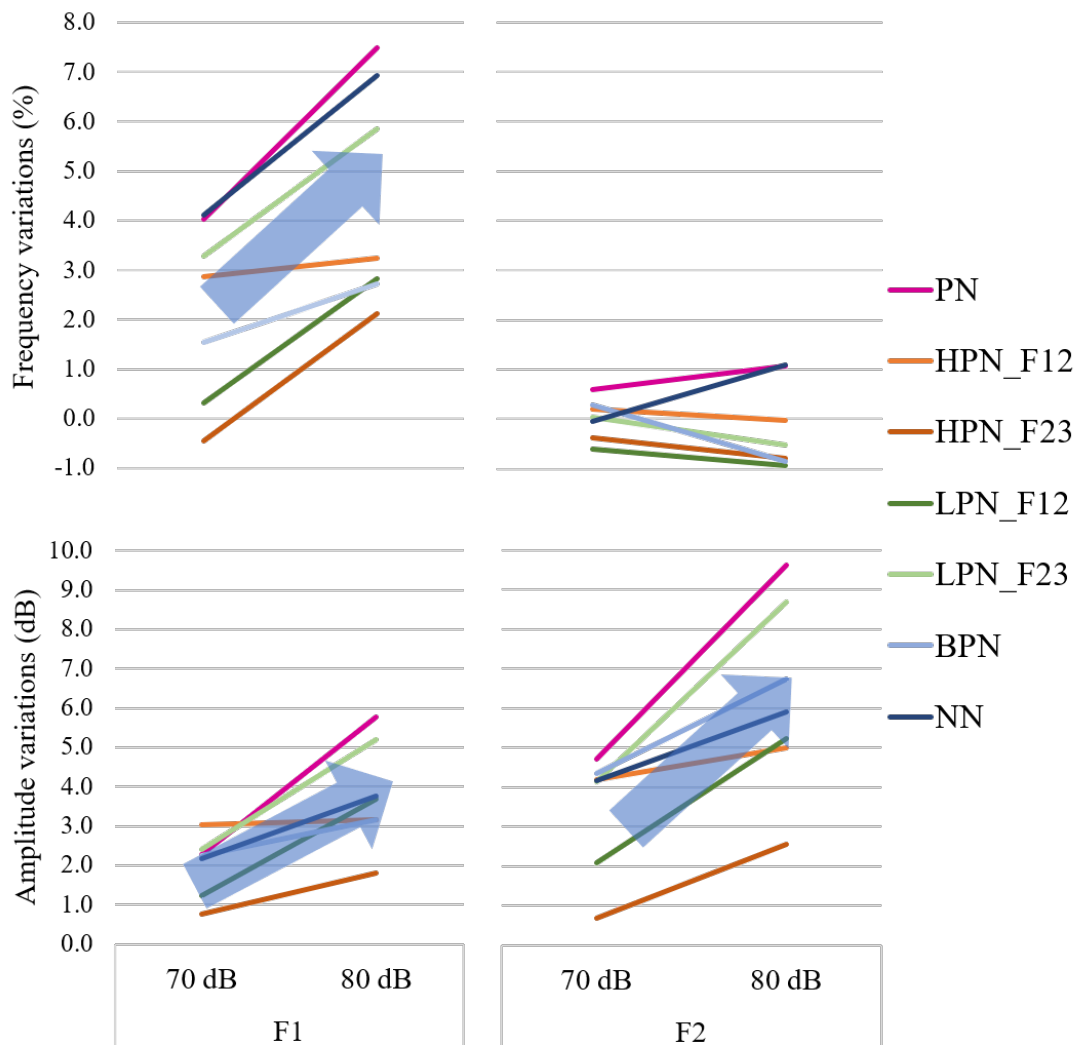


図 4.17: 雑音毎の音圧レベルによる F1 と F2 の周波数・振幅の変化

4.4 雑音環境における興奮パターン

4.4.1 結果

本研究では、ヒトの脳内に入るまでの、聴覚の特性を考慮するため、興奮パターン [32] を算出した。被験者 1 の /a/ におけるそれぞれの実験刺激の興奮パターンを図 4.18–24 に示す。縦軸を Excitation level とし、横軸を聴覚フィルタの中心周波数とした。図中の F1・F2 は、被験者 1 の /a/ の算出された F1・F2 を示している。PN の F1・F2 への影響は、他のノイズと比較して大きいことが分かる。また、図 4.24 から、F2 への影響は少なからず存在するということが分かる。また、BPN においては、F1 への影響は、聴覚上でもほとんどないということが分かる。物理スペクトルにおいては、F2 への影響はない NN や HPN_F23, LP_F12 であるが、聴覚におけるスペクトル形状を観測すると、F2 への影響が存在することが分かる。また、HPN_F23 の F2 への影響は、LPN_F12 と比べて小さいことが分かる。聴覚の特性を考慮すると、低域の周波数が高域へと影響を及ぼしやすいということが分かる。

4.4.2 考察

4.4.1 の結果から、聴覚上のスペクトルで雑音を観測すると、低域の周波数が高域へと影響を及ぼしやすいということがわかる。そのため、今回の発話変動の観察をするにあたって、発話に重要な低域に大きなパワーをもつ雑音であればあるほど、発話に大きな影響を与えていることが推察される。ヒトがこの雑音を知覚し、脳内で運動指令を出すということは、生成側にも影響が大いにあることが推察される。逆に、HPN のような低域側に大きな影響をもたないような雑音は、発話に重要な周波数である低域側への影響は小さいため、脳内の音声生成への運動指令に対しても影響が小さいことが推察される。

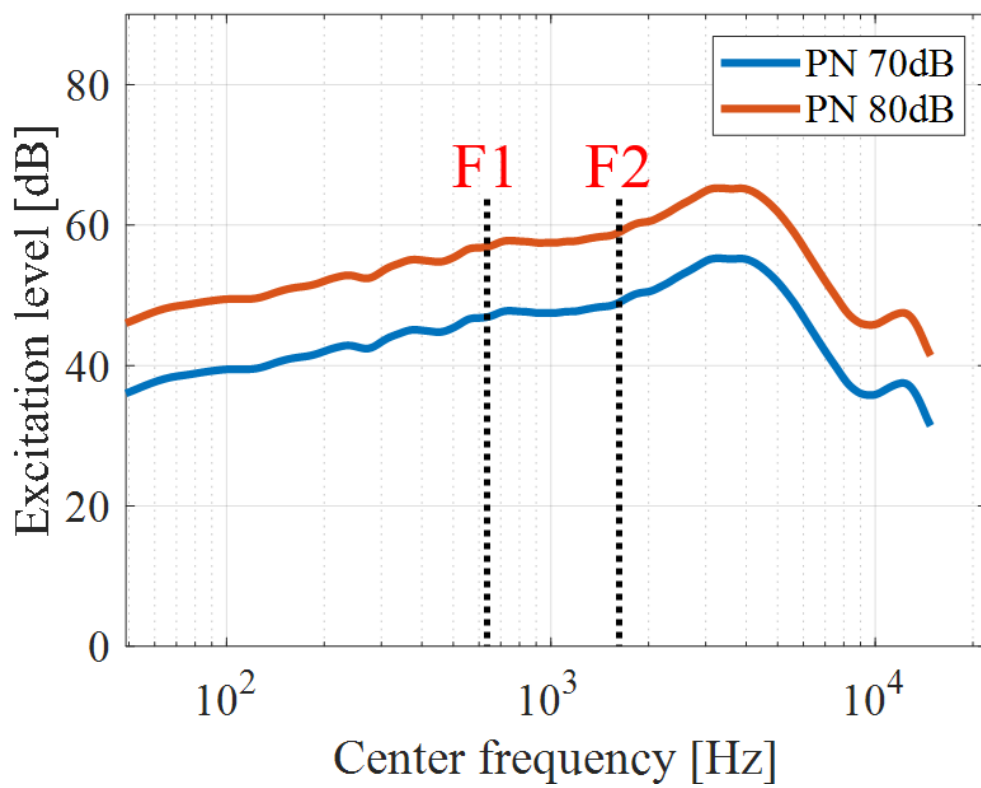


図 4.18: 背景雑音として PN を付加した際の興奮パターン (被験者 1,/a/)

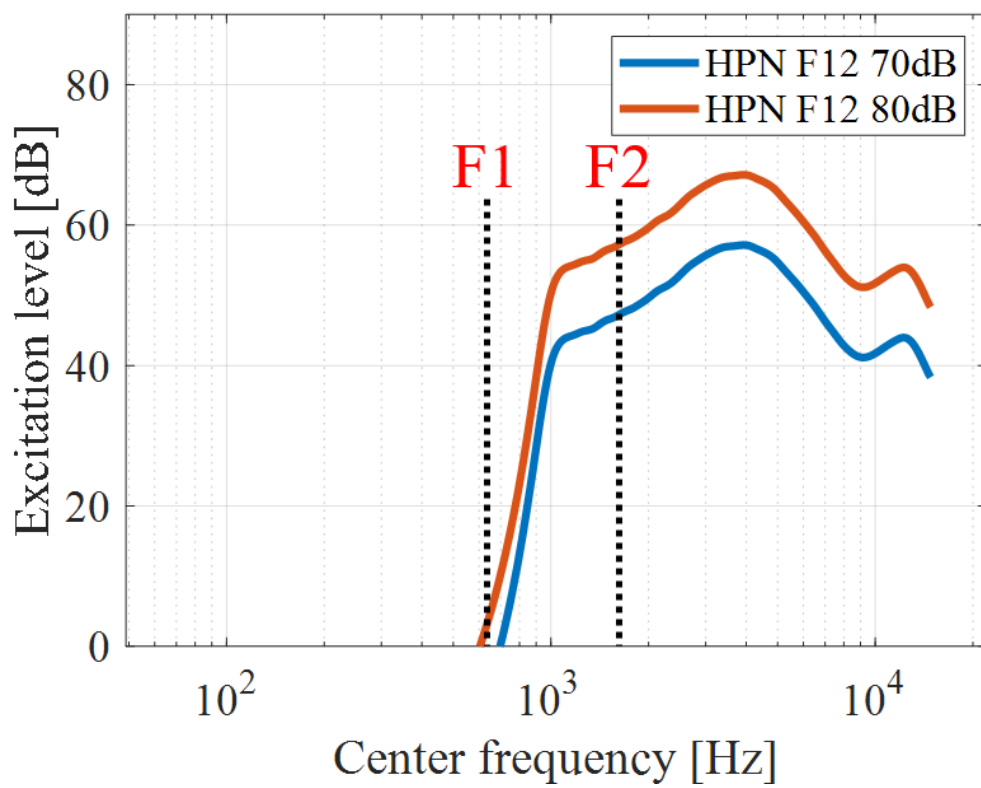


図 4.19: 背景雑音として HPN_F12 を付加した際の興奮パターン (被験者 1,/a/)

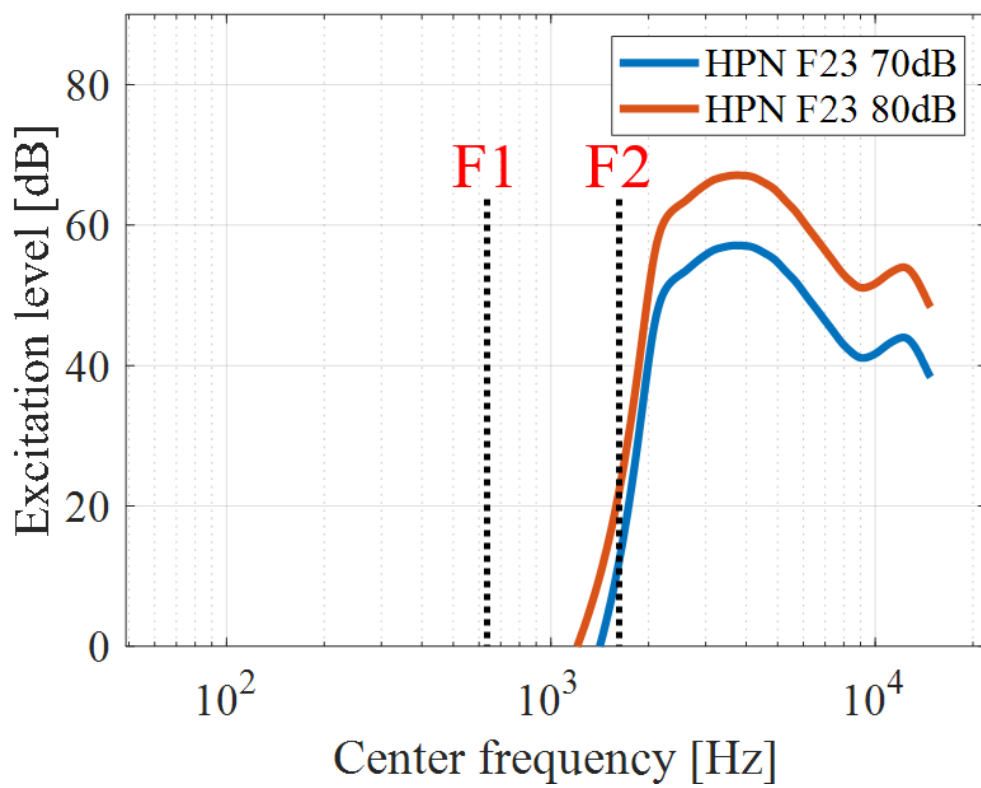


図 4.20: 背景雑音として HPN_F23 を付加した際の興奮パターン (被験者 1,/a/)

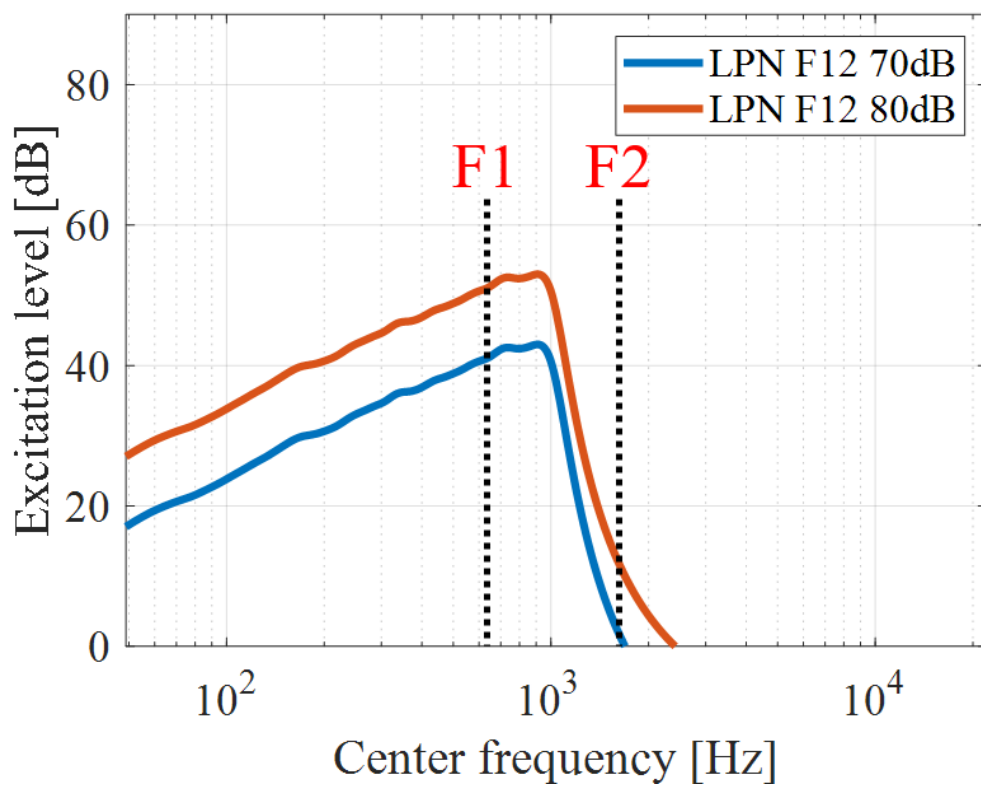


図 4.21: 背景雑音として LPN_F12 を付加した際の興奮パターン (被験者 1,/a/)

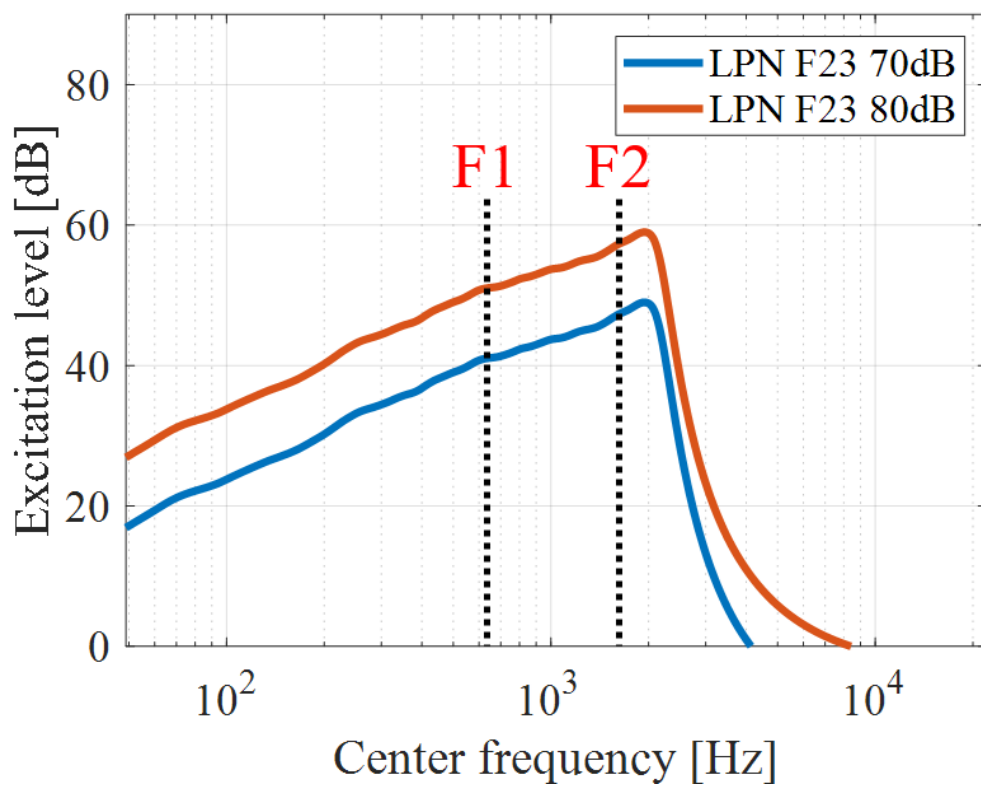


図 4.22: 背景雑音として LPN_F23 を付加した際の興奮パターン (被験者 1,/a/)

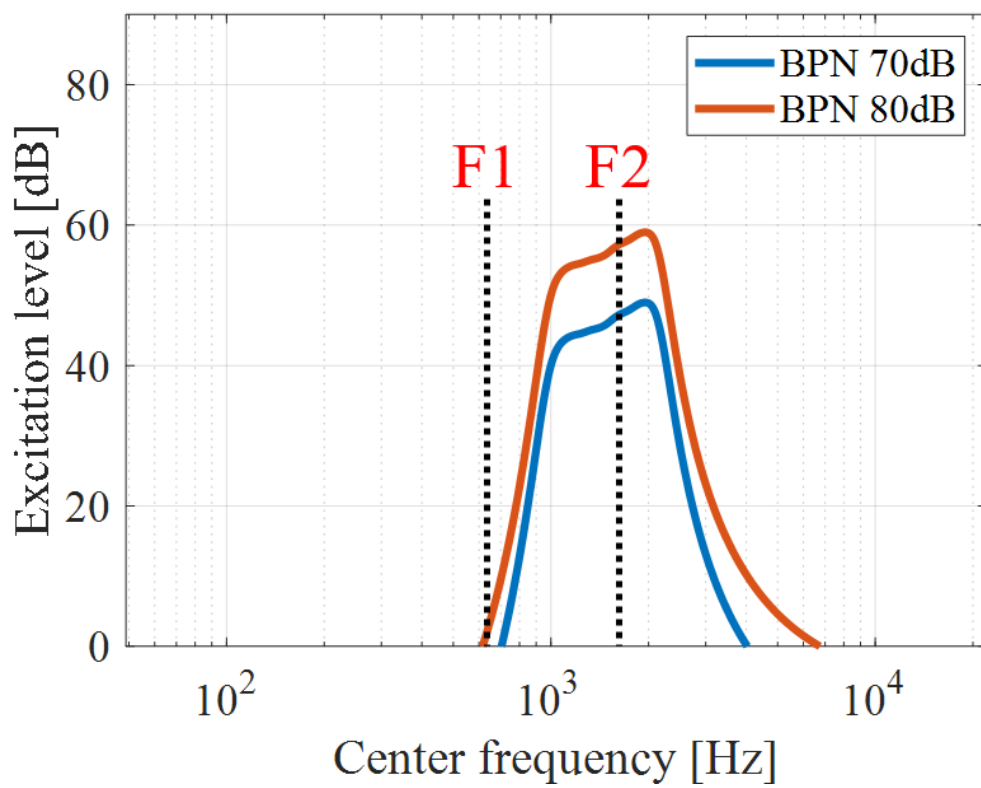


図 4.23: 背景雑音として BPN を付加した際の興奮パターン (被験者 1,/a/)

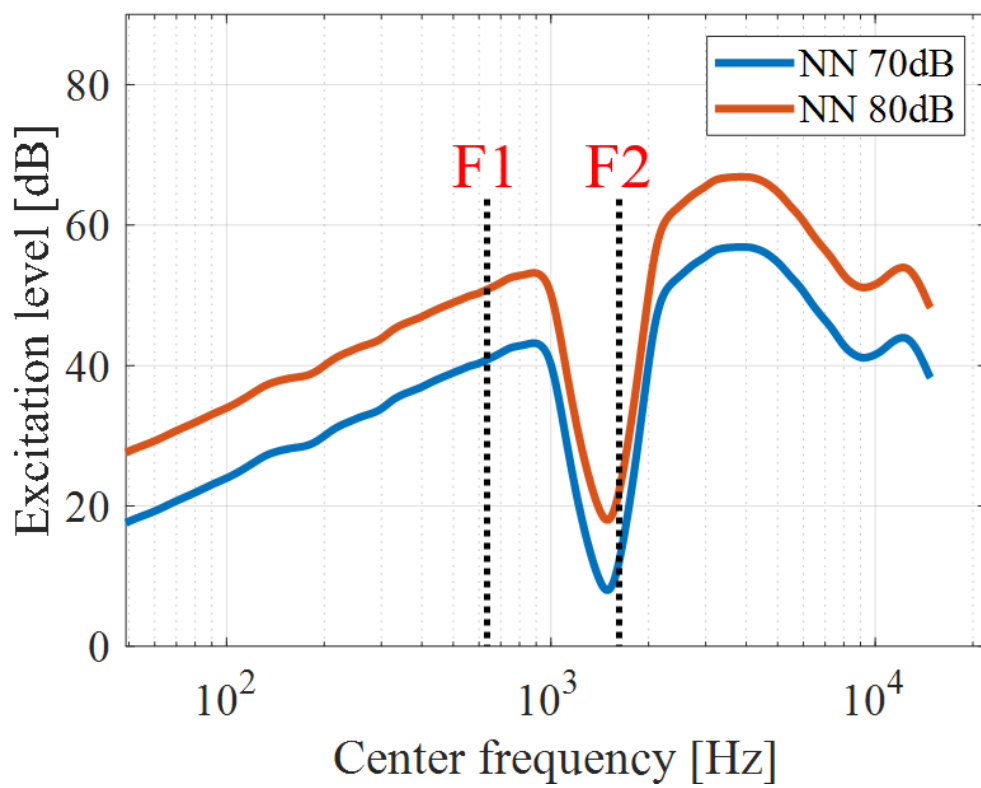


図 4.24: 背景雑音として NN を付加した際の興奮パターン (被験者 1,/a/)

4.5 刺激が発話に与える影響とそれらの変化

4.5.1 結果

刺激が聴覚神経に与える影響と、それに対する発話変化の関係性を調査するため、4.3.1の結果と4.4.1の結果を刺激・フォルマント毎にプロットし、比較した。その結果を、図4.25–28に示す。刺激のフォルマントへの影響を刺激毎に棒グラフで表している。刺激毎のフォルマントの変化を折れ線グラフで示している。図4.25では、主軸は、刺激のF1へのエキサイテーションレベルを示している。第2軸は、F1の周波数の変化を示している。F1への影響が大きな刺激であるほど、周波数の変化が大きい傾向があるということが分かる。図4.26では、主軸は、F2への雑音のエキサイテーションレベルを示している。第2軸は、F2の周波数の変化を示している。4.3.1の結果からも分かるように、フォルマントへの影響が大きな刺激であればあるほど、変化が大きいということはないということが分かる。図4.27では、主軸は、刺激のF1へのエキサイテーションレベルを示している。第2軸は、F1の振幅の変化を示している。この結果から、刺激のF1への影響が大きければ大きいほど、F1の振幅の変化が大きい傾向があるということが分かる。図4.28では、主軸は、刺激のF2へのエキサイテーションレベルを示している。第2軸は、F2の振幅の変化を示している。この結果から、刺激のF2への影響が大きければ大きいほど、F2の振幅の変化が大きい傾向があるということが分かる。また、図4.27と図4.28の結果を比較すると、F2の振幅の変化は、F1の振幅の変化と比較して大きいことが分かる。

4.5.2 考察

4.5.1の結果から、聴覚上においてフォルマントへの影響が大きい刺激ほど、発話に変化しやすい傾向があるということが分かった。このことから、ヒトは、周囲の雑音のフォルマントへの影響が大きければ大きいほど、発話がマスクされていると感じ、その部分のフォルマントを変化させる傾向にあるということが推察される。また、このときの変化は、F1は主に周波数方向に、F2は振幅方向に変化させることによって、発話を雑音の中で際立たせようとするという推察される。

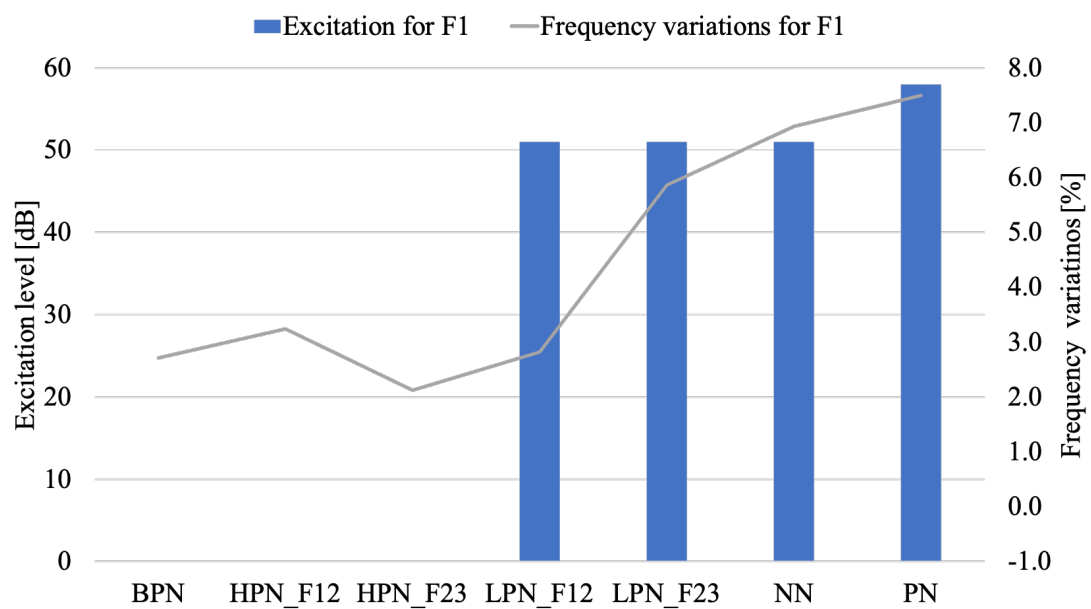


図 4.25: 刺激が F1 に与える影響と F1 の周波数の変化

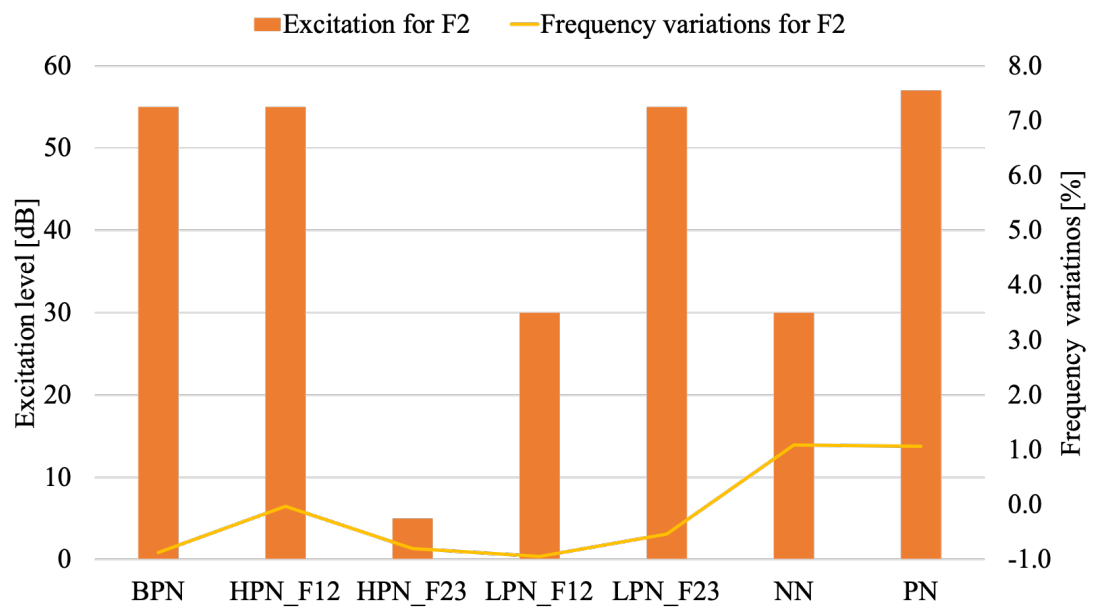


図 4.26: 刺激が F2 に与える影響と F2 の周波数の変化

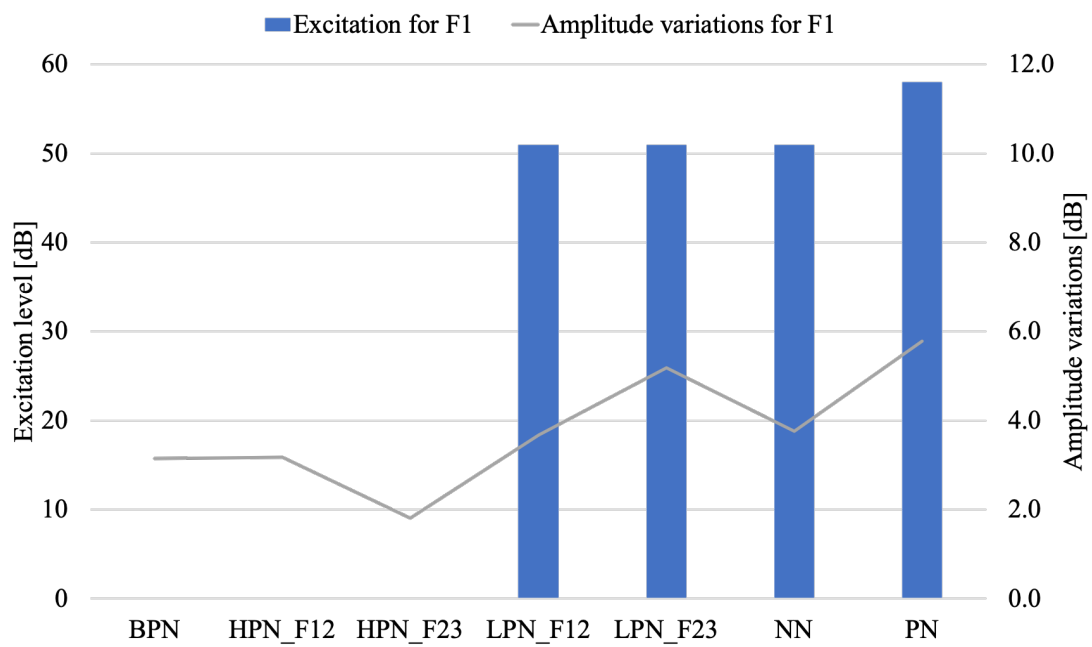


図 4.27: 刺激が F1 に与える影響と F1 の振幅の変化

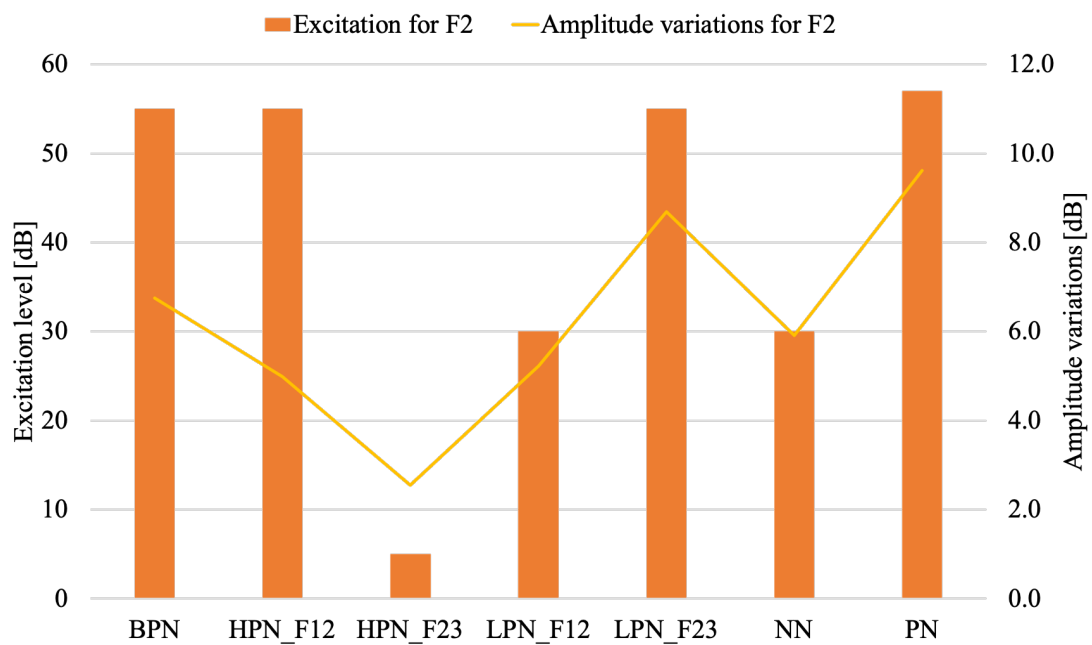


図 4.28: 刺激が F2 に与える影響と F2 の振幅の変化

第5章 総合考察

5.1 雑音による発話の変化

雑音環境下における母音発話時の話者の方略を調査し傾向をつかむために、音声のF1・F2の変化を調査した。F1・F2は母音の発話に重要である。そのため、7つの実験刺激はF1・F2に影響があるように、それぞれ2つの音圧レベルで作成された。そして、これらの雑音環境下で発話を収録し、静かな環境での発話と比較した。図4.17におけるPNの結果から、F1の周波数が背景雑音の音圧レベルの上昇に伴って上昇していることが分かる。これは、先行研究と同様の結果であることが分かる [19]。そのため、この結果から、本実験においてロンバード効果が発生したことが分かる。

5.2 背景雑音の音圧レベルによる発話の変化

雑音環境下における母音発話のF1・F2の周波数と振幅の変化を調査するため、70 dB、80 dBの2つの音圧レベルを用意し、発話の変化を観測した。その結果、4.17においてF1は主に周波数を、F2は主に振幅方向に変化させるという傾向が得られた。この傾向は、今回実験に使用したすべての刺激において見られる傾向であった。先行研究において、実験刺激の種類によっては、F1・F2の周波数が実験刺激を避けるように下方にも遷移することもあるという推察がされていた [21]。しかし、今回の実験では、そういった傾向は観測されなかった。また、F1の周波数は、舌の上下方向の位置と、F2は舌の前後方向の位置と、相関があることが知られている。そのため、今回の実験から、雑音環境下において発話する際において、ヒトは、雑音の種類に関わらず、静かな環境で発話する場合と比べて、舌の位置を下方に遷移させ、音声の強度を高くするということが推察される。今回の実験から、舌の前後方向の位置に関して、実験刺激の音圧レベルによる傾向は無かった。

5.3 背景雑音の種類による発話の変化

雑音の種類による発話の変化を調査するため、7種類の刺激を用意し、刺激毎の母音発話のフォルマント周波数と振幅を調査した。図4.17より、これらの変化は

刺激の種類に依存することが分かった。その中でも、F1・F2 への影響が大きい刺激であればあるほど、4.3.1 で発見された発話の変化が大きいことが分かった。図 4.17 の LPN と HPN の結果から、背景雑音として LPN を付加した際の発話の方が、HPN を付加した際と比較して音圧レベルの差による発話の変化が大きいことが分かった。また、HPN_F23 の発話の変化は今回使用された実験刺激の中で最も変化が小さい。これは、母音発話時において、背景雑音として F1 および F2 に影響のある低周波数の雑音がある場合は、それ以外の雑音がある場合と比べて母音発話が変化しやすいということが推察される。図 4.2 から、PN の母音発話への影響は他の実験刺激と比べて大きいことが分かる。PN は他の実験刺激と比べて低域が強調された刺激である。また、分析された図 4.24 の結果から、この刺激は F2 にも影響がある。また、図 4.17 の PN と NN の結果から、音圧レベルによる F1 の周波数の変化は類似していることが分かる。しかし、背景雑音として NN を付加した際の F1 および F2 の振幅の変化は、PN を付加した際の振幅の変化よりも小さかった。この違いは、刺激の F1・F2 の影響に依存すると推察される。

第6章 結論

6.1 明らかにしたこと

本研究では、母音の発話に重要なF1・F2に焦点を当て、そこに対する影響の違う実験刺激を7種類用意し、それらの刺激が背景にある時、ヒトはどのような方略をとって発話を変化されるのかについて調査した。その結果、刺激の種類に関係なく、その音圧レベルによって、F1は周波数方向に、F2は振幅方向に主に変化する傾向があることを明らかにした。また、刺激毎で見れば、興奮パターン上で、F1・F2への影響が大きい低周波のであればあるほど、発見した傾向の変化が大きいことが分かった。これは、ヒトの発話メカニズムの解明につながる知見である。本研究で得られた知見は、災害が起こった際に流れる避難誘導音声への応用 [33] や、雑音環境下における自然で明瞭度の高いアナウンス音声などへの応用が期待される。

6.2 のこされた課題

本研究では、雑音の種類およびその音圧レベルによって、雑音環境下における母音発話の静的な変化傾向を調査した。本研究で得られた知見をもとに、雑音環境下における動的な変化傾向を調査することで、ヒトの雑音環境下における発話変化のメカニズムの解明によりつなげることが可能であると考え。また、実際に音声コミュニケーションを行う際には、母音だけではない。そのため、ヒトの雑音環境下における音声コミュニケーションのメカニズムを解明するためには、子音や、それらを組み合わせた単語などについても調査する必要があると考える。

謝辞

本研究を進めるにあたり，多大なるご指導ならびにご鞭撻を賜りました赤木正人教授に心から感謝いたします。また，研究室会議をはじめ，様々な機会で助言をいただきました鵜木祐史教授に感謝いたします。日ごろ，研究や私生活の悩み・疑問に対して親身に相談に乗ってくださった，赤木・鵜木研究室のメンバーの皆様に心より感謝いたします。本学での研究生生活を支え，見守っていただいた両親に心から感謝いたします。また，研究や，就職活動についての悩みについて親身に相談に乗ってくださった，赤木・鵜木研究室のOBの皆様にも感謝いたします。最後に，本学で同甘共苦した同期の皆様にも心から感謝いたします。

研究業績

国際会議

1. Tomoya Nishigaki, and Masato Akagi, “Influence of auditory feedback on uttering vowel speech in noisy environment,” 2020 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP’20), 2020.

国内発表

1. 西垣 朋哉, 赤木 正人, “雑音環境下における聴覚フィードバックが母音の発話に与える影響,” 日本音響学会聴覚研究会資料, vol. 49, No. 7, pp. 493–498, 千葉, 2019年11月.

参考文献

- [1] 廣谷 定男, “「聞くと話す」の脳科学,” コロナ社, 2017.
- [2] J. W. Black, “The effect of delayed side-tone upon vocal rate and intensity,” *J. Speech Disorders*, 16, 1, pp. 56–60, 1951.
- [3] B. S. Lee, “Artificial stutter,” *J. Speech Disorders*, 16, 1, pp. 53–55, 1951.
- [4] E. Lombard, “Le signe de l’ élévation de la voix, *Ann. Mal. De L, Oreille et du Larynx*, vol. 37, pp. 101–119, 1911.
- [5] J. J. Dreher, and J. O’Neill. “Effects of ambient noise on speaker intelligibility for words and phrases,” *The Journal of the Acoustical Society of America*, pp. 1320–1323, 1957.
- [6] A. L. Pittman, and T. L. Wiley. “Recognition of speech produced in noise,” *Journal of Speech, Language, and Hearing Research*, 2001.
- [7] 北原 鉄朗, 小暮 計貴, 永 真宏, 鈴木 光, “騒音下における声の張り上げ現象の計算機による実現に向けて,” *人口知能学会研究会*, pp. 33–37, 2015.
- [8] 野村 行弘, 呂 建明, 関屋 大雄, 谷萩 隆嗣, “雑音量に依存しない音声/雑音領域判別法を利用した音声強調の改良,” *日本音響学会誌*, vol. 62, no. 1, 2006.
- [9] 山田 安紀子, 今泉 敏, 原田 勇彦, 細井 裕司, “感音性聴覚障害者における時間的要因を変化させた母音の識別,” *Audiology Japan*, vol. 32, pp. 130–136, 1989.
- [10] M. Cooke, S. King, M. Garnier, V. Aubanel, “The listening talker A review of human and algorithmic context-induced modifications of speech,” *Comput. Speech Lang*, vol. 28, no. 2, pp. 543–571, 2014.
- [11] J. C. Junqua, “The lombard reflex and its role on human listeners and automatic speech recognizers,” *J. Acoust. Soc. Am.*, pp. 510–524, 1993.
- [12] H. Brumm, H. Slabbekoorn, “Acoustic communication in noise,” *Advances in the Study of Behavior*, vol. 35, no. 9, pp. 151–209, 1991.

- [13] A. R. Bradlow, and J. A. Alexander, “Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners,” *The Journal of the Acoustical Society of America*, pp. 2339–2349, 2007.
- [14] W. V. Summers, D. B. Pison, R. H. Bernacki, R. I. Pedlow, , and M. A. Stokes, “Effects of noise on speech production,” *Acoustic and perceptual analyses. The Journal of the Acoustical Society of America*, pp. 917–928, 1988.
- [15] M. Garnier, N. Henrich, D. Dubois, “Influence of sound immersion and communicative interaction on the Lombard effect,” *J. Speech Lang. Hear. Res.* 53, pp.588–608, 2010.
- [16] Y. Uemura, M. Morise, and T. Nishiura, “The Lombard speech recognition based on the voice conversion towards neutral speech,” *ICA2010, PaperID, 167*, 2010.
- [17] 程島 奈緒, 荒井 隆行, 栗栖 清浩, “雑音・残響下における発話の音響的特徴の話者変動,” *IEICE Technical Report*, vol. 69, no. 11, 2009.
- [18] C. Davis, J. Kim, K. Grauwinkel, and H. Mixdorff. “Lombard speech: Auditory (a), visual (v) and av effects,” In *Proceedings of the Third International Conference on Speech Prosody*, pp. 248–252, 2006.
- [19] T. V. Ngo, R. Kubo, D. Morikawa, and M. Akagi, “Acoustical analyses of tendencies of intelligibility in lombard speech with different background noise levels,” *Journal of Signal Processing*, pp. 171–174, 2017.
- [20] L. M. Stowe, and E. J. Golob, “Evidence that the Lombard effect is frequency-specific in humans,” *J. Acoust. Soc. Am.*, vol. 134, no. 1, pp. 640–674, 2013.
- [21] S. Matsumoto, M. Akagi, “Variation of Formant Amplitude and Frequencies in Vowel Spectrum uttered under Various Noisy Environments,” *NCSP2019, Honolulu, Hawaii, USA, Mar.*, pp. 4–7, 2019.
- [22] T. Chiba, M. Kajiwara, “The vowel: Its nature and structure” *Iwanamisyoten, Tokyo*, 2003.
- [23] 河原 英紀, “聴覚フィードバックの発声への影響-ヒトは自分の話声を聞いているのか?-,” *日本音響学会誌*, vol. 59, no. 11, pp. 670–675, 2003.
- [24] 鏑木 時彦, “音声生成の計算モデルと可視化,” *コロナ社*, 2010.
- [25] 粕谷 英樹, 鈴木 久喜, 城戸 健一, “年齢, 性別による日本語 5 母音のピッチ周波数とホルマント周波数の変化,” *日本音響学会誌*, vol. 24, no. 6, pp. 355–364, 1968.

- [26] 日本音響学会編, “音響用語辞典,” コロナ社, 1988.
- [27] 松岡 理絵, “聴覚系と発話系の相互作用に関する研究,” 北陸先端科学技術大学院大学 修士論文, 2005.
- [28] Acoustic core: <http://www.acousticcore.com/>, (2020年2月現在)
- [29] B. R. Glasberg, B. C. J. Moore, “Prediction of absolute thresholds and equal-loudness contours using a modified loudness model,” *The Journal of the Acoustical Society of America*, vol. 120, no. 2, pp. 585–588, 1991.
- [30] B. C. J. Moore, R. Glasberg, “Suggested formulae for calculating auditory-filter bandwidths and excitation patterns,” *The Journal of the Acoustical Society of America*, vol. 74, no. 3, pp. 750–753, 1983.
- [31] Z. Chen, G. Hu, B. R. Glasberg, B. C. J. Moore, “A new method of calculating auditory excitation patterns and loudness for steady sounds,” *Hearing Research*, vol. 282, pp. 204–215, 2011.
- [32] ISO 532-2: “Methods for calculating loudness - Part 2: Moore-Glasberg method,” *Acoustics*, 2017.
- [33] 赤木 正人, 鷓木 祐史, 久保 理恵子, 小林 まおり, “「災害時に必要な情報を音声により確実に伝える」インテリジェント避難誘導音声呈示システムの研究開発,” *信学技報*, EA2017-8, pp. 45–50, 2017.