人工知能学会研究会資料
JSAI Technical Report
SIG-Challenge-B303-06

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

Towards Computational Implementation of Phenomenon of Raising Voice in Noisy Environment

北原 鉄朗 † **小暮 計貴** ‡ **吉永 眞宏** † **鈴木 光** †

Tetsuro Kitahara† **Kazuki Kogure**‡ **Masahiro Yoshinaga**† **Hikaru Suzuki**†

† 日本大学文理学部 ‡ 日本大学大学院総合基礎科学研究科

†**College of Humanities and Sciences, Nihon University**

‡**Graduate School of Integrated Basic Sciences, Nihon University**

{kitahara, kogure, yoshinaga, hikaru}@kthrlab.jp

Abstract

雑音の大きい環境では，人間は自然と声を張り上げてしまうことがある．このことは，人間による音声発話には聴覚系からのフィードバックが有することを示唆しており，雑音の大きい環境で対話相手に確実に聴こえる発話をするのに役立っていると思われる．本研究では，雑音の大きい環境で有効に動作する音声対話システムを実現する上で，この現象を計算機上で再現することが鍵になると考え，そのための課題と予備的検討について述べる．

1 はじめに

ヘッドフォンをして音楽を聴いている状態で話しかけられると，妙に大きな声で返事をしてしまう場合がある．これは，音声発話に聴覚系のフィードバックが強く働いていることを示している．ヘッドフォンに限らず，周囲の雑音が大きい環境にいると，自然と声を張り上げてしまうことはよく知られている．これにより，静かな場所で発声された音声に比べてインテンシティが大きくなるだけでなく，基本周波数やフォルマント周波数が高くなるなど，様々な音響的特徴が変化する．このことは**ロンバード効果** [Lane 71] と言われている．

一方，音声対話システムにおける音声発話部には，このような特徴はもちろんない．周囲の雑音状況とは無関係に，あらかじめ決められた声色で音声を合成し，あらかじめ決められた音量でそれを再生する．そのため，雑音の状況が動的に変わるような環境では，周囲が静かなときには声が大きすぎ，うるさいときには逆に聞こえないという事態になりかねない．携帯電話のようなユーザが個人的に所有・使用するような場合はユーザが自ら音量調整することもできるが，駅での運行案内など，公共の場で用いられることを想定したシステムでは，ユーザが音量を調整するのは容易ではない．

音声対話システムが広く社会で用いられるようになる上で，雑音耐性が重要であることは言うまでもない．これまで雑音下音声認識については非常に多くの研究がなされてきたが，雑音の状況が動的に変化する環境で，システムの発話を確実にユーザに聴こえるようにする工夫については，あまり研究されてこなかった．音声強調や音声明瞭化などの研究は様々なものが存在する（e.g., [Arai 02, 荒井 07, 竹山 06]）が，雑音が動的に変化する環境で，音量やその他の音響的特徴を自動的に調整して，ユーザが確実に発話内容を聞き取れるようにする試みではなかった．

我々は，このようなことの実現を目指す上で，上述のロンバード効果が参考になると考えている．つまり，ロンバード効果を計算機上で再現することが，動的に変化する雑音状況に適切に対処する音声発話への近道だと考えている．本稿では，ロンバード効果について簡単にまとめた後，それを計算機上で実現する上での課題について述べる．その後，できるだけ単純化して実現した場合の予備的な検討結果について述べる．最後に，その検討結果によって分かった問題点を挙げ，その解決案について議論する．

2 ロンバード効果について

ロンバード効果については様々な研究結果があるが，ここではその一例として程島らによる研究結果 [程島 09] を紹介する．

程島らは，静かな環境（Q），雑音のある環境（N），2種類の残響のある環境（R1，R2）で，東京方言話者 4 名（男女 2 名ずつ，22~37 歳）に様々な単語や音素バランス文を発声してもらった．雑音は白色雑音を使用し，発話者の耳元で平均 80dBA になるように騒音計を用いて音量を調整した．

その結果，基本周波数（F0）と第 1 フォルマント（F1）については，Q 条件に比べて N 条件，R1 条件，R2 条件

いずれも有意に上昇した．一方，第 2 フォルマント（F2）については，N 条件，R1 条件，R2 条件いずれも Q 条件に比べて有意差はなかった．子音と母音のインテンシティ比（CVR）は，N 条件，R1 条件，R2 条件いずれも Q 条件に比べて減少した．音圧レベルは，Q 条件と比較して N 条件，R1 条件，R2 条件いずれも増加したが，Q 条件に対する増加量は R1 条件，R2 条件の方が N 条件よりも少なかった．

3 ロンバード効果を計算機上で実現する上での課題

ここでは，このロンバード効果を計算機上で実現する上で解決すべき課題について述べる．課題は，大きく次の 2 つに分けることができる．

課題 1 雑音測定

課題 2 発話パラメータ設定

課題 1 は，その名の通り，どのように雑音の音量を測定するかである．課題 2 は，次の 3 つに細分化される．

課題 2-1 どの音響的特徴を変化させるか．

課題 2-2 どのタイミングで音響的特徴を変化させるか．

課題 2-3 どの程度の値に音響的特徴を変化させるか．

課題 1 に対して最も単純な方法が

案 A 音声対話（ユーザ音声の入力）用以外に雑音測定用のマイクロフォン（あるいはマイクロフォンアレイ）を用意し，それで計測された音響信号の振幅を求める

という方法である．この方法は単純で実装も容易であるが，自己発話（システム自身が発話した音声をこう呼ぶこととする）やユーザ発話も雑音とみなしてしまう場合がある．そのため，システムの判断によって自己発話の音量を大きくすると，それによって雑音が大きくなったと判断されるので，より一層自己発話の音量を大きくしようとし，発散してしまうという問題がある．この問題を解決しようとするのが次の案である．

案 B 自己発話やユーザ発話を抑制してから振幅を求める

自己発話については，雑音に重畳される自己発話の音響信号は既知なので，その分を減算して抑制することで，雑音のみの音量をより正確に推定できると考えられる．また，ユーザ発話については音声対話（ユーザ音声入力）用のマイクロフォンから得られた音響信号を参照信号として同様の処理を行う方法が考えられる．

課題 2-1 については，

案 A 音量のみ制御する

案 B 基本周波数，フォルマントも制御する

の 2 案が考えられる．案 A の場合は，音声合成エンジンによって生成された合成音声の再生系を制御すればよいので，音声合成エンジンと実装を切り離すことができ，比較的容易に実装できるというメリットがある．案 B の場合は，音声合成エンジンに対して基本周波数やフォルマントを制御する必要があるので，そのような制御が可能な音声合成エンジンを使用する必要がある．

課題 2-2 については，

案 A 音声発話開始時にのみパラメータ設定を変更する

案 B 音声発話中も時々刻々と動的にパラメータ設定を変更する

の 2 案が考えられる．案 A は実装が単純化されるだけでなく，上で述べたような，システムによる発話の音量が上がることによって雑音の音量が上がったと判定されてシステム発話の音量を上げてしまい，これが繰り返すことによって音量の設定が発散する事態を防ぐことができる．しかし，発話開始直後に大きな雑音が突発に発生しても対処できないという問題がある．そのため，長い発話には特に不向きである．

課題 2-3 は，たとえば雑音が 80dBA だと分かったときに，システム発話の音量やその他の音響的特徴をどれだけ上げ下げしたら，容易に聞き取れて大きすぎない音声になるか，という課題である．最も単純な方法は，

案 A 雑音の音圧レベルをいろいろ変えてみて，各音圧レベルに対してちょうどいい音量設定（音量以外も変えるならそのパラメータ）を実験的に調査する

という方法であろう．この案の最大の問題点は，環境依存になってしまうことである．音の聴こえ方は雑音の種類，対話音声用のスピーカーの設置角度など，様々な要因によって変化してしまうため，運用環境ごとに調査が必要となってしまい，汎用性に問題が生じる．また，調整すべきパラメータが増えたときに調査は大変困難になる．そこで，次のような案が考えられる．

案 B 音声の明瞭度を何らかの基準で定義し，その基準を満たすようにパラメータを最適化する

システム発話の音響信号は既知であるので，その音響信号が他の音源に比べて十分に優勢であるかを何らかの方法で測定できれば，その優勢度を音声の明瞭度として用い，この値が一定値を超えるように音量やその他のパラメータを自動的に最適化することができるであろう．たとえば，システム発話の音響信号のスペクトルピークとそれ以外（雑音）のスペクトルピークを比較し，SN 比を算出してこれを明瞭度とみなすなどの方法が考えられよう．

案 C 音声認識させてみて一定以上の精度が出るようにパラメータを最適化する

ユーザが音声を聞き取れるか（内容を認識できるか）が重要であると考えるのであれば，雑音入りの音声をシステムで認識させてみて，その認識が成功するように音量などを調整するという方法も考えられる．しかし，システムによる音声認識の精度は人間によるそれに比べて（特に雑音環境下では）低く，人間がきちんと聞き取れれば十分という観点では，システムによる音声認識精度を基準とするのは，過剰要求であるとの考え方もあるであろう．

このように，ロンバード効果を計算機上で実現するには，様々な課題がある．我々は現在，これらの課題を解決すべく，検討を進めている．まずは最も単純な方法（各々の案 A）を試し，その後，より複雑な方法（案 B，案 C など）を試すという方針で進めている．

4 予備的検討

本章では，3. の議論に基づいて行った予備的検討 [鈴木 14] について述べる．この予備的検討では，次の方針を採用した．

課題 1 案 A を採用.

課題 2-1 案 A を採用.

課題 2-2 案 B を採用.

課題 2-3 案 A を採用.

課題 2-2 のみ案 B を採用したのは，この予備的検討に先だって行った実験で案 A を採用したところ，発話開始直後に大きな雑音が発生してシステム発話が聞き取れない事態が頻出したためである．

4.1 システム構成

実験用システムの構成を図 1 に示す．この実験用システムは，利用者が発話用マイクロフォンの手前に位置して音声対話を行うことを想定している．発話用マイクロフォンの近くにシステムによる音声発話用のスピーカーが設置され，発話用マイクロフォンとは別に，雑音測定用にマイクロフォンアレイが設置されている．音声対話の内容は東京都内の乗り換え案内とし，利用者は「○○駅から▲▲駅まで行きたい」のように発話を行うと，システムは「駅すぱあと Web API」を用いて最短経路を取得し，音声合成による発話を行う．ただし，今回の実験用システムでは，システムによる発話が聞き取れるかどうかのみに目的を限定し，発話用マイクロフォンは使用しないものとする．また，後述のように，実際の経路を探索して案内するのではなく，あらかじめ用意した音声を聞かせるものとする．

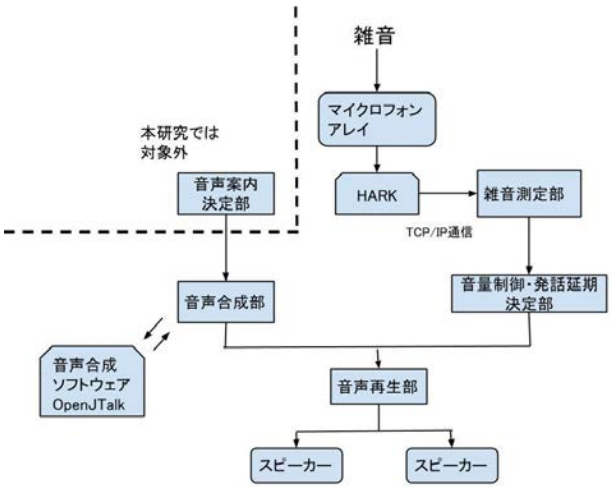

図 1: システム構成図

4.1.1 雑音レベルの計測

雑音レベルは，マイクロフォンアレイから得られる音響信号に基づいて推定する．現在の実装では，7ch のマイクロフォンアレイ「Microcone」からロボット聴覚オープンソフトウェア「HARK」[奥乃 10] を利用して約 1 秒間毎に音響信号を取得する．それに対して RMS を計算し，あらかじめ騒音計を用いて作成した RMS と騒音レベル (dB) の変換式に代入し，騒音レベル (dB) を算出する．

4.1.2 再生音量の変更

音量の変更は計算した雑音レベルを元に行う．システムの発話音量より周囲の雑音の方が大きい場合，雑音と同じ値まで発話音量を増幅する．また，周囲の雑音がシステムの発話音量より小さい場合は発話音量の縮小も行う．これにより環境に最適な発話音量の自動調整を実現する．

4.1.3 発話の延期

音量調整による雑音対策の他に発話の延期による対策も施す．これは，電車の警笛など最大音量を超える突発的な雑音に対応するためである．現在の実装では，64dB を超える雑音を感知した場合は発話を中断し，1 秒毎に雑音の計測を行い，64dB を下回ったときに発話を行うようになっている．

4.2 実験

提案手法によって利用者がシステムの発話を聞き取りやすくなったかどうかを実験する．

4.2.1 実験方法

実験は外からの騒音が入りにくい部屋で行った．被験者は 21 歳から 24 歳の正常な聴力を有する男性 3 人，女性 3 人の計 6 人である．被験者の位置を中心に 60 度おきに 6 箇所にスピーカーを設置した（図 2）．以下の流れで実験を行った．

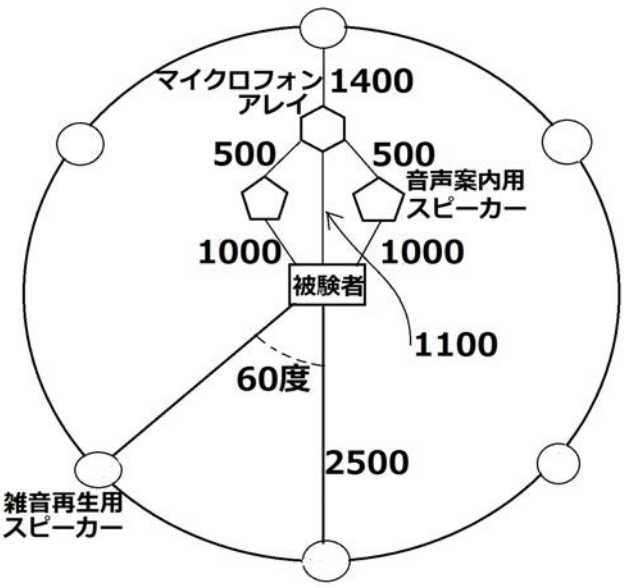

図 2: 実験時の機器配置図 [mm]

1. スピーカーから雑音を再生する.
2. 雑音に慣れてもらう (30 秒間).
3. 音量調整を行わない音声発話を再生.
4. 提案手法の音声発話を再生.

雑音は Microcone を使って東京都内の駅のホームで録音したものを再生し，発話には OpenJTalk で音声合成した女声を用いた．発話頻度は 10 秒に 1 回である．被験者は実験中に音声発話が聞こえたかどうかを 6 段階で評価する．それぞれの評価は以下のようにした.

1 まったく聞こえなかった

2 声は聞こえるが何を言っているのかわからない

3 聞き取れるが，大部分が聞き取りにくい

4 聞き取れるが，一部聞き取りにくい

5 聞き取れる

6 聞き取れるが，音量が大きすぎる

発話内容は,「○○から△△までの料金は□□円です.」とし，次のような 4 つの選択肢から聞き取ったものを選んでもらった.

- 「青砥から青井までの料金は 230 円です.」
- 「青砥から青井までの料金は 230 円です.」
- 「青井から青砥までの料金は 530 円です.」

表 1: 通常の音声発話による結果．各セルの左は各回答番号を選んだ回数，右はその中で選択問題に正答した割合を表す.

被験者	回答番号 6		5		4		3		2		1	
A	1	100%	10	90%	5	80%	3	67%	1	—	0	—
B	0	—	17	94%	2	100%	2	50%	0	—	0	—
C	0	—	12	100%	5	100%	3	100%	0	—	0	—
D	0	—	14	93%	1	100%	2	50%	4	—	0	—
E	0	—	16	94%	1	0%	0	—	0	—	3	—
F	0	—	15	93%	2	0%	2	0%	2	—	0	—
平均	0.2	100%	14	94%	2.7	75%	2	58%	1.2	—	0.5	—

表 2: 提案手法による音声発話の結果．各セルの左は各回答番号を選んだ回数，右はその中で選択問題に正答した割合を表す.

被験者	回答番号 6		5		4		3		2		1	
A	1	100%	15	93%	5	0%	0	—	0	—	0	—
B	0	—	16	94%	5	100%	0	—	0	—	0	—
C	0	—	19	100%	2	100%	0	—	0	—	0	—
D	0	—	16	100%	4	100%	0	—	1	—	0	—
E	0	—	21	100%	0	—	0	—	0	—	0	—
F	0	—	20	95%	0	—	0	—	0	—	0	—
平均	0.2	100%	17.8	97%	2.7	69%	0	—	0.2	—	0	—

- 「青井から青砥までの料金は 530 円です.」

音声発話を聞き取れるかどうかを確かめるのが目的なのでこの実験では「駅すぱあと Web API」を用いず，また，元々料金を知っていることの効果を防ぐため，でたらめな料金を案内することとした．他に「音量は適切だったか」,「発話の遅延は適切だったか」など設問回答型のアンケートも実施した.

4.2.2 実験結果と考察

通常の音声発話を再生した結果を表 1，提案手法による音声発話の結果を表 2 に示す.

表は横軸が聞き取りやすさの違いによる 6 段階の評価，縦軸が各被験者を表している．結果の左側がそれぞれの評価が記録された回数，右側が類似文章による選択問題の正答率となっている．それぞれの結果で，音声発話が正常に行われなかったデータは削除している．また，6 段階評価の [1], [2] については聞き取れなかった評価のため，4 択問題による聞き取り判断は行わないものとする.

表 1 の結果から，音量調整を行わない場合は 6 人中 4 人が [1] または [2] の評価をしているため，聞き取りに困難を感じていると判断できる．平均に着目すると，すべての人が 1 回以上は聞き取りに困難を感じている結果となった．また，評価 5 を選択した回数は提案手法を用いた場合，通常再生より平均で 3.8 回増加し，6 人中 5 人で評価 5 を選択した回数が増加している．類似文章による選

表 3: 再発話

被験者	発生発話数	評価 5	評価 4
A	6	5	1
B	6	5	1
C	6	6	0
D	7	6	1
E	7	7	0
F	6	5	1
平均	6.3	5.7	0.7

択問題の正答率も上昇していることから，提案手法によって聞き取りやすくなっていると言える．

６人の被験者のうち，被験者 B だけが提案手法を利用して適切な音量で聞き取れたと回答した回数が低下した．ここで被験者 B の類似文章による選択問題の正答率に注目すると聞き取れ方が不安定な時の場合，通常の音声発話の正解率が 50 %なのに対して提案手法の正答率は 100 %となった．これは聞き取りやすく感じた回数は低下したが，実際に正しく聞き取れた回数は増加したことを意味している．

また，聞き取りが不安定な時の正答率の平均が低下した原因については以下のようなことが考えられる． 提案手法が音量調整を行う際に周囲の環境音が静かなときに必要以上に再生音量を小さくしてしまう． 実際に，アンケートでは提案手法で周りが静かになったときに音量が音声発話の音量が小さくなりすぎていたとの回答が複数得られた．

提案手法による延期が発生した発話の全てが評価５または評価４であった．その結果を表３に示す．左の列から「被験者」,「延期が発生した発話数」,「延期が発生した発話で評価５が記録された回数」,「延期が発生した発話で評価４が記録された回数」となっている．通常の音声発話と提案手法を比較すると，評価５の出現回数が平均 3.8 回程度増加し，評価４以下の出現回数が平均 3.5 回減少していることから，発話の延期が行われることによって聞き取りづらく感じていた発話が聞き取りやすくなったと考えられる．

5　今後の改良に向けて—まとめに代えて

4. で述べた予備的検討では，自己発話が雑音と判断されて自己発話の音量上昇によって発話の音量を上げ続けてしまう現象を防ぐ根本的な解決は行わなかったため，雑音が 64dB を超える場合は音声発話を行わず，雑音が収まるまで発話を延期させる方策をとった．これにより，電車の通過のような突発的な雑音が発生した場合でも発話を聞き取れるようになったが，一方，待たされる時間が長いという意見があった．このことから，3. での議論の通り，課題 1 に対して案 A では不十分で，案 B を検討することが重要であることが明らかになった．課題 2-3 に対しては，今回 1ヶ所でしか実験を行っていないので，設置環境に対する汎用性については未検証であるが，今回取った方法（案 A：音圧レベルごとにちょうどいい音量などの設定値を実験的に調査する）は事前調査に要する手間が大きく，案 B あるいは案 C を検討する必要があることが分かった．課題 2-2 に対しては，音声発話開始後に突発的な雑音が発生したときに，すぐにそれに合わせて音量が上昇する点はよかったが，雑音の音量変化に合わせてシステム発話の音量を細かく上下させたことにより，発話が不自然になることがあった．また，発話途中に雑音が収まると，それに合わせて音量が低下するために，聞き取りにくくなる場合があって．そのため，音量を上昇させる場合は素早く，低下させる場合はゆっくりと行うなどの工夫が必要であることが分かった．

このように，3. で述べた課題の解決が重要であることが 4. の予備的検討で明らかになった．雑音環境下で有効に働く音声対話システムの実現のため，3. での議論に従って研究を進めていきたい．

謝辞

本研究は，SCAT 研究助成による助成を受けて実施されたものである．また,「駅すぱあと Web API」をご提供くださった（株）ヴァル研究所に感謝する．

参考文献

[Arai 02] Arai, T., Kinoshita, K., Hodoshima, N., Kusumoto, A., and Kitamura, T.: Effects of Suppressing Steady-state Portions of Speech Intellibility in Reverbant Environments, *Acoust. Sci & Tech.*, Vol. 23, No. 4 (2002)

[Lane 71] Lane, H. and Tranel, B.: The Lombard Sign and the Role of Hearing in Speech, *J. Speech Hear. Res.*, Vol. 14, pp. 677–709 (1971)

[奥乃 10] 奥乃 博：ロボット聴覚の現状と展望, 日本ロボット学会誌, Vol. 28, No. 1, pp. 2–5 (2010)

[荒井 07] 荒井 隆行：音声に関するバリアフリー, 音響研資, H-2007-66, pp. 377–382 (2007)

[竹山 06] 竹山 佳成：騒音環境下における車室内発話音声の分析とその合成に関する研究, Master's thesis, 北陸先端科学技術大学院大学 (2006)

[程島 09] 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, 信学技報, SP2009-69, pp. 43–48 (2009)

[鈴木 14] 鈴木 光, 吉永 眞宏, 小暮 計貴, 北原 鉄朗：雑音環境下のための音声案内システム：周囲の雑音レベルに合わせた音量の自動調整, 情処全大, 6S-1 (2014)